United Nations ECE/CES/2018/2

Economic and Social Council

Distr.: General
27 March 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixty-sixth plenary session
Geneva, 18–20 June 2018
Item 4 (b) of the provisional agenda
Data integration for measuring migration

Guidance on data integration for measuring migration

Note by the Task Force on data integration for measuring migration

Summary

The document presents a short version of the *Guidance on data integration for measuring migration*. The Guidance was prepared by a Task Force composed of Spain (Chair), Austria, Italy, New Zealand, Switzerland, United Kingdom, United States, Eurostat and UNECE.

The current short version of the Guidance is prepared for translation purposes and it excludes the country case studies. The Guidance includes: a discussion of the definition of data integration, with reference in particular to international migration statistics (Chapter II); a discussion of the results of the survey on data integration practices (Chapter III); summary results of the thirteen national case studies (Chapter IV); a discussion and list of metadata (Chapter V), the conclusions (Chapter VI) and a number of general recommendations for countries (Chapter VII).

The full text of the Guidance (including the country case studies) has been sent to all members of the Conference of European Statisticians for electronic consultation. It is available at: http://www.unece.org/index.php?id=47411. Subject to a positive outcome of the consultation, the CES plenary session will be invited to endorse the Guidance.

GE.18-04790(E)

1804790

Please recycle

Contents

Chapter		*Paragraphs*	*Page*
I.	Introduction	1–10	3
II.	Definition of 'data integration'	11–31	4
	A. General features	11–23	4
	B. Working definition	24–31	6
III.	Survey on data integration practices	32–47	8
	A. Overview of main results of survey	35–40	8
	B. Recommendations	41–42	9
	C. Summary of results	43–47	9
IV.	Case studies	48–49	10
	A. Summary of the case studies	50–70	10
V.	Metadata	71–77	12
VI.	Conclusions	78–88	14
VII.	Recommendations	89–98	16
	A. Improve access to administrative data for national statistical offices	90	16
	B. Use administrative data for migration statistics	91–92	16
	C. Combine data from different sources using 'presence signals'	93	16
	D. Pay attention to the quality of integrated data	94	16
	E. Be transparent about data integration methods used and develop standards	95	17
	F. Promote international comparison and exchange of migration data	96–98	17
VII.	Further work	99–103	17
	A. The potential of big data	100	17
	B. Longitudinal data	101–103	18
References			18

I. Introduction

1. Data integration has been a topic of great interest in recent years, especially as data needs have grown under constrained fiscal climates, combined with increased respondent burden and privacy concerns. The UNECE High-level Group on Modernisation of Official Statistics (HLG-MOS) had a project on data integration in 2016–2017. The project included experiments with data integration in different areas and led to developing a practical online guide to data integration for official statistics.[1] Eurostat has supported research on methods to improve data integration, such as the ESSnet project in the area of Integration of Survey and Administrative Data.[2] This project was an initial attempt to create a common methodological basis for integrating different data sources.

2. However, many countries have an interest in data integration specifically for the purposes of international migration measurement. Given the challenges in collecting migration statistics, it is often useful to collect data from several different sources. Data integration can reduce coverage or accuracy problems that may concern overall migration stock and flow data, and enhance the richness of migration data by adding socio-demographic or economic dimensions to existing data.

3. At the European level, the importance of data integration to improve migration statistics was reaffirmed on the occasion of the 2017 annual conference of NSI Directors General, which included a statistical session on 'Population Movements and Integration Issues - Migration Statistics'.[3] The 'Budapest Memorandum' discussed at the conference and approved by the European Statistical System Committee includes among the action points: *"To support the identification, assessment and adoption of new methods and data sources, particularly the increased use for statistical purposes of administrative data sources of appropriate quality ensured through ongoing quality assessment – either single registers, linked data from several administrative sources or combined with survey sources, and the opportunities offered by new data sources (e.g. Big Data)"*

4. Methods to integrate different administrative data sources within a country (e.g. population registers, health, work, taxation, or education records) to supplement information missing from various sources (e.g. to measure outmigration for those who fail to deregister) have proved successful in some countries. In other cases non-administrative data sources can provide information missing from administrative sources (e.g. use of household surveys to collect information not included in registers). 'Big data' from utility and phone companies are also being considered in some countries as supplementary information. Other uses of data integration could be in reconciling different migration figures as derived from different sources, particularly in estimates for 'hard-to-count' migration populations, such as irregular migrants or emigrants.

5. Noting the importance and timeliness of this topic, in 2014, the UNECE-Eurostat Work Session on Migration Statistics discussed the utilization of different administrative data sources to measure international migration. The discussion focused on the opportunities and challenges faced when using administrative data, particularly given the differing levels of development of administrative data systems across the region. Ways to improve cooperation between national migration services, statistical agencies, agencies in charge of registers, and other producers of administrative data were considered as were methods of integrating various administrative data sources to improve measurement of migration. Participants also emphasised the need for practical advice and guidance on the

[1] https://statswiki.unece.org/display/DI/Data+Integration+Home

[2] https://ec.europa.eu/eurostat/cros/content/data-integration_en

[3] 103rd DGINS Conference (Budapest, 20–21 September 2017). Papers and presentations are available at http://www.ksh.hu/dgins2017/presentations.html

production of metadata to facilitate comparisons between migration estimates produced by different countries.

6. The Work Session concluded by recommending further methodological work on the topic of integration of multiple data sources for measuring international migration, including data sources within a country and between different countries, and good practices in communication between national statistical offices and producers of administrative data.

7. Based on these discussions, the Conference of European Statisticians (CES)[4] Bureau established the Task Force on Data Integration for Measuring Migration in October 2015, to prepare a set of guidelines and description of good practices for integrating different data sources to improve the measurement of immigration, emigration and net migration. This publication presents the results of the work of the Task Force.

8. The publication provides a general overview of this subject as well as an overview of the types of data integration that are already in use in various countries – whether disseminated through regularly published data or as part of pilot studies. Principles of best practices based on this overview are also provided to serve as guidelines for improving data integration for measuring migration in different countries. Country experiences are documented in this publication on the basis of a survey of migration data providers in nearly 50 countries as well as on more detailed case studies for several countries.

9. Data integration is addressed in this publication by examining both *macro-data integration* – the comparison/statistical modelling based on data, which are aggregates (statistics) of individual level records – as well as *micro-data integration* – the integration of data based on linkage/matching of individual level records. Differing levels of overlaps of variables and/or individuals between different sources are also analysed.

10. Further details on the above concepts are provided as part of the conceptual background and definition of data integration given in Chapter II. Chapter III provides an analysis of the results of the survey of migration data providers. Practical examples of data integration in migration are shown through in-depth case studies from a number of countries. The summary result of the case studies are presented in Chapter IV (the case studies are available in the full version of this document, in English only). A proposal on metadata to be considered with regard to international migration statistics is presented in Chapter V. Finally, the conclusions and general recommendations based on the survey and the case studies are provided in Chapter VI.

II. Definition of 'data integration'

A. General features

11. The measurement of migration has since long proved to be a very challenging task. Although in recent time considerable efforts have been made in both the academic and official statistics domains to improve the quality of migration statistics, there are still

[4] The Conference of European Statisticians is composed of national statistical organizations in the UNECE region (for UNECE member countries, see www.unece.org/oes/nutshell/member_states_representatives.html) and includes in addition Australia, Brazil, Chile, China, Colombia, Japan, Mexico, Mongolia, New Zealand and Republic of Korea. The major international organizations active in statistics in the UNECE region also participate in the work, such as the statistical office of the European Commission (Eurostat), the Organization for Economic Cooperation and Development (OECD), the Interstate Statistical Committee of the Commonwealth of the Independent States (CIS-STAT), the International Labour Organization (ILO), the International Monetary Fund (IMF), the World Trade Organization (WTO), and the World Bank.

relevant margins of uncertainty as for their accuracy. More recently, in a wider statistical context, the attention has been drawn on the opportunities offered by 'data integration' as potential additional action for the improvement of statistics (e.g., Eurostat, 2009, 2013; ESSnet, 2008, 2011, 2014; UNECE-HLG MOS 2017).

12. 'Data integration' is often mentioned as one of most effective approaches for enriching and improving the statistical information. Various activities and projects over the last years dealt with data integration, and adopted different definitions. For instance, according to the Generic Statistical Business Process Model (GSBPM), data integration is an activity in the statistical business process when data from one or more sources are integrated. In the HLG-MOS project mentioned above, data integration is defined as "the activity when at least two different sources of data are combined into a dataset".[5] Yet, to the best of our knowledge, there is not an internationally agreed definition of data integration in the statistical community. There are various features which may help to characterize data integration, such as number and typology of data sources, methodology, timeliness, response burden, and not least its purpose.

13. Intuitively, integrating requires the availability of at least two inputs. These inputs are the datasets, i.e. organized information derived from selected data sources by means of a statistical operation. In the case of migration, datasets can be classified in four categories: (datasets derived from) statistical surveys (including exhaustive surveys, i.e. censuses), administrative registers, big data and geographical information. These typologies can be integrated in all possible combinations and the process can be repeated, generating datasets of mixed nature which can be in turn integrated with other datasets.

14. The way datasets are integrated mainly depends upon whether the information they contain is on micro or macro level and, for individual records, upon the availability of a common identifier. For the integration at micro level, this latter feature makes the difference in the choice between statistical matching and record linking, the latter usually applied when an identifier is available in all datasets being integrated. In the integration at macro level, i.e. between data aggregated from single records, the process should take place in two steps: first, cleaning from differences in concepts and operational definitions; second, reconciling ('balancing') the data, using various statistical techniques or mere expert opinion.

15. In the case of migration data, the most common individual identifier is the Personal Identification Number (PIN). In fact, micro-level integration is preferred for the setup of population registers, of which migration statistics are often a by-product. It should be noted that a PIN is not the only possible common identifier, as different identification coding can be used to link data related to sub-national aggregations, administrative entities, file records, and so on.

16. The data sources used as input to a population register can be many and in some countries additional data sources have been integrated over time, enriching the statistical information on the population and, indirectly, on migration. This can be the case also for registers of the foreign population / aliens registers, although usually to a slighter extent. Data integration can thus be seen as an expanding process, where to the first integrated datasets, others are added following lessons learnt, refinement of integration methodologies, and access to new data sources.

17. Another distinctive feature of migration measurement is the fact that, given the nature of its phenomenon shared by two countries (either receiving and sending countries, or hosting country and country of birth, or whatever the variable for identifying migrants), data sources other than national can be used. The exploitation of this intrinsic international feature is mostly limited to the exchange of aggregated data, given the high concerns about

[5] https://statswiki.unece.org/pages/viewpage.action?pageId=169018059

privacy and national security that countries may oppose any request of international migration microdata exchange. Examples of this latter approach do exist in the Nordic countries, where international microdata exchange is used to improve the quality of the population registers.

18. Adding new data sources may lead to changes in the timeliness of the statistical output, i.e. changes in the delay between the reference period and the availability of results. Whilst the general rule would be that the timeliness of a specific output from integrated data is given by the data source with the larger delay, statistical modelling may actually reverse this situation by giving the opportunity to release more timely estimates. It is likely that a worsening of the timeliness would only be accepted when the data integration brings a relevant improvement of the coverage or accuracy of the statistical measurement.

19. The reduction of response burden is sometimes mentioned as one of the opportunities offered by data integration. To date, most use of alternate data sources for migration statistics is achieved by replacing a data source by another data source with a lower burden on the respondents/data providers or by improving the statistical operation applied on the data source. From a conceptual perspective, this may be seen rather as an enhancement of the efficacy than an enrichment of the data availability, and therefore not exactly peculiar/pertinent to data integration.

20. Inputs from alternatives data sources may be used also for the sake of validating the statistical output of a single, official data source. In this case, it can be considered as a matter of *comparison* rather than *integration*. This may also be the first step in a process of progressive integration of these data sources, especially when the additional data sources do not comply with the requirements of official statistics.

21. Another case is when each data source covers a specific population (migrant) sub-group. The pooling of these data to produce an overall statistical output could be labelled as 'compilation' rather than pure 'integration', as there is no actual merging at the level of the single record (either individuals or aggregated entities).

22. From the considerations above, it derives that the qualifying feature of data integration is certainly the use of multiple datasets, but conditionally to the way these are used jointly. In fact, there are statistical operations which are somehow border cases with data integration, such as data compilation and data comparison. As for the other features, the level of detail of the data matters for the methodology to be applied, rather than for the identification of a data integration activity. Timeliness cannot be used either as identification criterion, given that data integration may in principle lead to an improvement as well as a worsening of the delay of the results, and that timeliness improvements can be gained as well by applying statistical operations other than integration. Likewise, any generic reference to quality improvement of the statistical output (including reduction of response burden) would not help, given that any statistical operation should in principle aim to such an effect. Obviously, this does not imply that data integration does not have such positive effects, but only that those references are not useful to identify distinctive features of data integration.

23. The outcome of a data integration activity should be an 'enriched' or 'higher quality' dataset. Thinking in terms of a generic data matrix $n \times p$, with n records and p variables, such enrichment could take place in both directions, improving the coverage (i.e., increasing n) and/or the information on the same records (i.e., adding new variables to the p set). As for the higher quality, this would not be reflected in changes of the size of the resulting dataset, but rather in its content (e.g., in the weights of a sample survey).

B. Working definition

24. An available definition of data integration is from the IT environment (SDMX, 2009) and it is the following: *"The process of combining data from two or more sources to*

produce statistical outputs". In the same reference, it is also clarified that *"Data integration can be at the micro-level, where it is often referred to as matching, or at the macro-level"*.

25. The definition above is based on inputs (*"two or more sources"*) and purpose (*"to produce statistical outputs"*). The latter is rather generic, because a statistical output can be seen as the end of any statistical process and it does not necessarily require data integration.

26. For the purposes of this document, the following working definition is hereby proposed: *"Data integration is a statistical activity on two or more datasets resulting in a single enlarged and or higher quality dataset"*.

27. Data integration can be processed at two main levels of aggregation: micro- and macro-data integration[6], defined as follows:

(a) 'Micro-data integration': the integration of data based on record linkage/statistical matching of individual level records using key identifying variables; and

(b) 'Macro-data integration': the combination of data based on aggregates (statistics) of individual level records.

28. The reference to a generic 'statistical activity' leaves it open the use of any methodology, either on macro or micro level, including expert opinion. It covers as well the work on conceptual differences, a paramount step in any statistical process. The explicit reference to statistical matching and/or record linking, although peculiar to data integration, would have excluded the macro-level integration.

29. *"Two or more datasets"* highlights the multidimensional nature of data integration. It also indicates that, whilst integration ends in a single outcome, it is an activity that needs to be repeated each time such outcome is targeted. In other words, for the regular production of statistics from more than one dataset, integration is not an occasional, once for all, statistical activity. The 'integrated' result is generated each time a fresher data input is available.[7]

30. The use of the word 'datasets' instead of 'data sources' points to the fact that subject of integration are actual data, and not the data sources from which they are derived. For instance, merging two existing sample surveys in a single encompassing survey by modifying the questionnaires, the sampling design, etc., is in this context not considered 'data' integration. It also supports the view expressed just above that the integration does not transform the input data sources in new, mixed data sources: it is the outputs of those data sources that are the core of the activity. The integration operation is thus likely to be as systematic as the production of statistics from the data on which it is applied.

31. The outcome of data integration is first and foremost a "single" dataset: a new set of data where all the information from the input datasets is re-organized in a harmonized fashion. This is not simply the union of multiple datasets, as redundancies, conceptual differences, and any other factor of bias are supposedly properly treated. This is an "enlarged" dataset, meaning a structured new set of data whose dimensions cannot be smaller than the largest corresponding dimensions of the individual input datasets, and/or an "higher quality" dataset, whose elements have been changed due to data integration with a (possibly measurable) gain in the statistical quality.

[6] In some contexts the terms 'micro-macro' is used when micro and macro level data are combined. This approach can be considered here as a case of macro-data integration.

[7] Identifying conceptual differences in a context of data integration is possibly an exception to the 'repetition rule', because such activity would need to be carried out only once, unless the conceptual framework of the input data changes over time.

III. Survey on data integration practices

32. In order to collect information needed to carry out its work, the task force carried out a survey on data integration practices. A specific questionnaire was designed to collect information on current practices in the use of combined or integrated data sources for the measurement of immigration and emigration flows, and for outputting statistics of the migrant population and the foreign-born population.

33. The questionnaire was forwarded in September 2016 to National Statistical Offices of UNECE member countries[8]. Fifty-six countries provided responses to parts of the survey that had relevance to their data collection practice.

34. The results of the survey were analysed between the end of 2016 and the first quarter of 2017. The main results are summarized in the following paragraphs.

A. Overview of main results of survey

35. Some form of data integration is essential for enabling statistical production of international migration flows. Where possible, data integration is done at the individual record level but this is mainly achievable for countries with complete population registers that account for exits of nationals and entries of foreign-nationals.

36. For many countries the actual migration event is not registered by a single source. Statistics of migration flows can only be derived by merging contributing components that account for changes in the country's resident population due to migration.

37. Statistics on international migration flows using a single source are produced when it is possible to maintain an administrative collection of the migration event (actual border crossings) or when a total population register based system is available. However, quality assurance and further disaggregation of these statistics may require integration with supporting data sources. This may just be in the form of an adjustment of migration flows following a comparison with available statistics from other countries (mirror statistics). Grouping of individuals in a single source may be required for longitudinal observations of border movements as a way of classifying migrant status.

38. The survey results highlighted an extensive variety in the uses of data integration. In some countries population registers of nationals may not be linkable at the unit record level to a separate administrative collection on foreign nationals. In such cases the population register may serve as a source for cross-checking and imputation of missing data.

39. A population census or a population register would be the common sources for producing statistics on the foreign-born population. The population census may serve as a base population and subsequent integration with an administrative collection enables the updating of the foreign-born population statistics.

40. When multiple sources are informing the migration statistics some countries have reported on the need for a prioritization assessment of the sources in terms of timeliness, completeness, reliability, coherence and other quality dimensions.

[8] In addition to the 56 countries that are members of UNECE, the questionnaire was also addressed to other countries that participate regularly in the activities of the Conference of European Statisticians, including Australia, Chile, Colombia, Japan, Mexico and New Zealand.

B. Recommendations

41. As countries advance and standardize registers of social and population statistical datasets, data integration is more likely to take place at the unit record level across multiple sources. A greater attention to a quality assessment of the integrated data will necessarily become an integral part of the production of international migration statistics.

42. It is increasingly important for statistical offices to provide a transparent presentation of the rules and methods used for the data integration processes that produce the statistics of international migration flows and other types of international migration statistics. This opens up opportunities to develop methodological standards of data integration practices across the national statistical offices in the future. Individual countries will always have their own collection and data integration practice but there are likely to be more open technical forums for discussion of rules and methods used in the combining of sources for the production of migration statistics.

C. Summary of results

43. Nearly all respondent countries produce statistics on international migration flows and more than half of these countries would use a combination of data sources. Some countries would use a central register of all registrations and de-registrations of residents, and integration of the source at the individual record level with administrative collections of emigration and immigration flows, and deaths, supported a refined production of migration flows. Other countries would integrate population registers with survey information or other administrative collection of non-nationals at a macro-level.

44. Overall, about a third of the respondent countries (total of 35 responses) integrated the data sources at the macro-level using statistical modelling or other method to combine aggregated data, and a slightly less proportion of countries would apply integration methods at the unit record level. However, it should be noted that a few countries used other sources for cross-checking and complimentary information only. About half of respondent countries noted observation units from the integrated data sources were partially overlapping; a smaller proportion had identical observation units when combining sources; and a few reported the use of mutually exclusive observation units.

45. Less than half the countries prepare statistics on international migration flows from a single source. Statistics on migration flows from a single source can frequently be produced when it is possible to maintain an administrative collection of all border crossings or when migration flows can be derived from a population register. Another less frequent option is the use of the population census to indirectly estimate migration flows from population stock measures.

46. Around 90 percent of respondent countries produced statistics or collected data on the foreign born population. Just over a half of these countries would use more than one data source and usually by data integration at the unit record or macro level. The main types and features of the sources used, either single or integrated, were a 5-yearly population census, population register, survey, or administrative collection. The main purpose of integration was reported to be for production of statistics on the foreign born population.

47. Around 60 percent of respondent countries reported they did not produce other statistics on migrants or the foreign-born population. So a few countries reported that integration of data sources allowed the preparation of statistics such as asylum seekers and grants, residence permits, work permit holders, irregular foreign workers, socio-economic characteristics of migrants, and reason for settlement.

IV. Case studies

48. Taking into account the information collected in the survey on data integration practices, the task force identified a number of countries with significant data integration experiences. These countries were invited to provide a short text describing the primary migration data sources, the reasons why data integration is considered necessary, the methods used and the benefits of integration.

49. The texts provided by the countries are presented in the full version of the present document (available in English only) as a collection of significant practical examples of data integration on migration.

A. Summary of the case studies

50. The case studies show a broad range of situations that differ widely with regard to the sources available for migration statistics, the limitations of those sources, and the data integration approaches adopted. The examples provided by the case studies should not be considered as fully representative sample of the data integration approaches developed in the various countries. Moreover, when evaluating the experiences of the various countries, it should be kept in mind that national contexts may differ significantly with regard to a number of important aspects, and that a data integration approach adopted successfully in one country is not necessarily the best solution for other countries. Notwithstanding these limitations, the analysis of the case studies presented allows drawing some key points that could be useful for other countries.

51. With regard to *the countries with a population register*, in several cases (including Austria, Italy, Latvia, and Spain*) this register is the main source of migration data*. In these countries data integration is carried out mainly to improve the data quality. Data from other registers can be used, among other purposes, to adjust for missing de-registrations of emigrants or to identify deceased persons. In several countries data integration is also used to provide additional information on variables not included in the primary sources.

52. Several countries adopt practices based on the *'presence signals'* (in some countries called 'signs of life') of individuals in the various registers. In *Austria*, for instance, the analysis of the presence signals in other registers is carried out to produce an estimate of those who are in the population register but do not live anymore in the country. Data from the Austrian Social Security are used to identify deceased persons, and from various other administrative sources to adjust for missing de-registrations of emigrants.

53. In *Italy*, where only persons who are legally resident in the country can be included in the population register, the analysis of presence signals in the various sources allows improving the international migration estimates by identifying persons who are in the population register but have probably emigrated, and those who immigrated but are not in the population registers.

54. In *Latvia*, the analysis of the presence signals is carried out using various registers, including those on revenue, social insurance, education, health, and employment. Emigration estimates are produced also using mirror statistics from destination countries. Moreover, the quality of migration statistics is evaluated using data from household surveys, including LFS and EU-SILC.

55. In *Spain* the national population register ('padrón') is managed by the national statistical institute INE, and all persons living in the country – irrespective of their legal situation – have to be registered. In connection with the 2011 census, the registered population was linked with many other administrative registers in order to identify persons whose presence in the country was doubtful and those who were dead. For doubtful records, data integration techniques based on administrative data and a sample survey

carried out in connection with the census allowed estimating the number of persons who were in the register but were not actually living in the country.

56. In some countries (including Hungary, Israel, Netherlands and Switzerland), migration statistics are produced using *systems where the population register is one of several data sources*, not necessarily the most important one. In *Hungary*, a relatively complex system of registers is used, where the various sources are complementary to each other. Data on immigration and emigration are produced by integrating data among the various sources, in particular the Office for Immigration and Asylum and the Central Population Register managed by the Ministry of the Interior. Data integration allows producing high quality migration statistics and provides additional information to that available from the primary sources.

57. In *Israel*, data on migration flows and stocks derive mainly from border control registrations, and are provided by the Population and Immigration Authority (which maintains also the population register) to the Central Bureau of Statistics. The data are combined with other administrative files (including tax files, education files, etc.) and with population census and other statistical surveys to provide estimates of grater quality and add variables from other sources related to the same individual or events.

58. In the *Netherlands*, migration statistics (like all social and demographic statistics) are based on the System of Social Statistical Databases, resulting from data integration among over fifty administrative registers. It is considered that virtually all the population living in the country is registered, with few exceptions. Concerning emigration, about one third of those leaving the country do not de-register. When the authorities find that someone is missing an investigation is conducted and, if not found, the person can be registered as emigrated to an unknown country.

59. In *Switzerland*, three main registers are relevant for migration data: 1) the local population registers maintained by the municipalities; 2) a federal register for legally resident foreigners holding a permit of stay; and 3) a special federal register for foreign diplomats, members of consulates, international organizations, their family members and similar categories. A modest form of data integration is carried out by matching data collected at the federal level and those collected at local level. Moreover, some variables available only in the register for legally resident foreigners are added to the local population register records pertaining to the same observations.

60. With regard to the *countries without a population register*, the main sources of data on migration include passenger cards filled at borders, data on passports, visas and stay permits. As for the countries with population registers, data integration is carried out using various sources to improve the data quality and obtain more comprehensive information. In particular, data integration may help improving the information for immigrants on the actual place and date of settlement, which often differ from those stated at the time of arrival.

61. In *Australia*, data on arrivals and departures are compiled by the Australian Bureau of Statistics based on information derived from various processing systems including passenger cards, passport and visa information. Data are linked using a personal identifier which allows ABS to also generate travellers' individual histories. Migrant data sets are also linked with census data and tax information.

62. In *Canada*, Statistics Canada receives administrative data on immigrants and non-permanent residents. Since the intended province and date of settlement not always corresponds with the actual place and date of settlement, data integration is carried out using other sources such as tax files. Record linkage is also carried out using data from other sources (such as the Canadian population census, the National Household Survey, the US Office of Migration Statistics) to provide estimates of great quality and obtain more comprehensive information on socio-economic characteristics of migrants.

63. In *New Zealand*, imaged passenger cards are linked to electronic border movements and passenger records to include additional information and provide validation. Moreover, for travellers who have indicated their intentions to stay in the country or leave it on a long-term basis, travel and passenger data are linked for the 16 months prior to the reference month, in order to confirm the status of New Zealand resident at the time of travel.

64. In the *United Kingdom* there are many sources of official statistics that provide information on different aspects of international migration flows and stocks. Concerning flows, the main source is the International Passenger Survey (IPS), a multi-purpose sample survey that collects information from passengers as they enter or leave the country. As the IPS does not take into account the changing intentions of passengers, in order to produce long-term international migration estimates data integration is carried out using data from IPS, LFS, Home Office immigration administrative data, and estimates of the flows at the land border of Northern Ireland (not covered by the IPS).

65. In the *United States* the Census Bureau produces annual estimates of international migration flows by adding various subcomponents derived from different data sources. For some subcomponents a limited integration (at the aggregate level) is carried out using some foreign and administrative data, including the American Community Survey, foreign censuses, and military movements from the Department of Defense. Integration of international migration data at the micro-level is not done at the Census Bureau, although some experiments were carried out.

66. To summarize, the case studies show that virtually all countries covered use data integration to improve the quality of migration statistics, and to provide additional information to that available from the primary sources.

67. In countries with a population register, data integration – also through the analysis of the 'presence signals' – may allow identifying: 1) persons who emigrated but are still in the population register; 2) persons who immigrated but are not in the population register, and 3) deceased persons.

68. In countries without a population register, data integration may help improving the information for immigrants on the actual place and date of settlement, which often differ from those stated at the time of arrival.

69. The case studies showed many examples of data integration conducted at micro or macro level. For example, in cases like Hungary or Canada different sources are micro-linked. This micro-integration is carried out by linking administrative sources on foreign people among them and with the population register to avoid overlapping and to provide additional variables. On the other hand, an interesting example of macro-integration is found in the United Kingdom, where a border passenger survey is combined, by calibrating, with another source (Labour Force Survey) to allow territorial breakdown of incoming flows.

70. As expected, many different types of sources are used for data integration, including various administrative or statistical registers, other administrative data (including border control data), and statistical surveys (including censuses). Few countries use 'mirror statistics' from destination countries to improve their emigration estimates. This is done mainly at the aggregate level, although there are examples of international exchange of microdata, namely among the Nordic countries (not included in the case studies), or restricted to neighbour countries.

V. Metadata

71. When migration estimates produced by different countries are compared, it is very important to have access to comprehensive metadata, in order to assess in detail the comparability of the data.

72. The metadata can be ideally split in three parts: the first is specific to each data source, describing their main features and how their datasets are transformed before they are subject to integration; the second part looks at the way the datasets generated from the listed data sources are integrated; the third part would be based on some measures of quality, basically to prove its increased informative content due to integration. This information should be provided for each relevant migration statistics, i.e. for migrants stocks, migration flows, and any of their sub-groups of interest.

73. The first information that should be provided is the list of data sources used to produce the integrated data. In such a list, the following metadata could accompany each data source:

(a) Name;

(b) Typology (administrative register, sample survey, census, big data, etc.);

(c) Owner (including statistical authority of another country);

(d) Rules of access / data provision (if the owner is other than national NSI);

(e) Population of reference;

(f) Date/period of reference of the data;

(g) Frequency of updating;

(h) Timeliness (i.e., the period of time between the date of availability of the data and their date/period of reference);

(i) Level of detail of the original data (micro, macro);

(j) Level of detail of the dataset used for integration (micro, macro);

(k) Description of any method applied to transform the original dataset into an input suitable for data integration;

(l) Dimensions of the dataset (n x p);

(m) The variables contained in the data source (or broad description if they are many);

(n) The variables contained in the data sources that are retained for integration (possibly subset of the previous item);

(o) The variables used to integrate this data source with the other(s) (possibly subset of the previous item);

(p) Availability of the PIN (this is in principle included in the previous item, but given its relevance it can be useful to dedicate a specific space to it) and its main features whether relevant.

74. Assuming that the integration process is implemented sequentially, for each step, i.e., for each pairs of datasets being integrated, the following information could be provided:

(a) Step number;

(b) Datasets involved (using names or number from the list of individual data sources);

(c) Frequency of the integration procedure;

(d) Variables used for integration (if applicable);

(e) Description of the methodology for integrating the datasets;

(f) Share of data overlapping;

(g) Main issues / difficulties in the integration of these datasets, such as reporting of non-linkage, methods and results from estimation of false-positive link rates;

(h) Dimensions (n x p) of the resulting integrated dataset.

75. In some cases, and particularly when migration statistics are derived from population registers, the number of data sources may be relatively high, and thus the reporting of the metadata can become burdensome. In such cases, it may be suitable to prepare just once a general report on the functioning of the system. However, it should also be taken into account that it may be important for the users to get a clear understanding on how the final data are produced through a possibly complex migration processing system. Complexity should not prevent transparency.

76. The third part of the metadata, which requires much further work, can be developed around the following initial set of quantitative measures of quality, aiming to highlight the positive effect of data integration:

(a) Difference in the number of records / size of the covered population between the final integrated dataset and the smallest single dataset: $n^* - \min(n)$;

(b) Difference in the number of variables between the final integrated dataset and the smallest single dataset: $p^* - \min(p)$;

(c) Difference in the timeliness between the final integrated dataset and the less timely single dataset: $t^* - \max(t)$;

(d) Difference in the number of records / size of the covered population between the final integrated dataset and the largest single dataset: $n^* - \max(n)$;

(e) Difference in the number of variables between the final integrated dataset and the largest single dataset: $p^* - \max(p)$;

(f) Difference in the timeliness between the final integrated dataset and the most timely single dataset: $t^* - \min(t)$;

(g) Gain in the quality of estimates, such as percentage decrease of the variance of the estimates, or any quantitative measure of the improved accuracy.

77. The metadata can be enriched by a qualitative assessment of the impact of integration, such as feedbacks given for the improvement of the original data sources and reduction of response burden, possibly supported by specific references / examples, and by an overall conclusion about the integration process.

VI. Conclusion

78. Migration statistics are probably the most difficult element in the field of social statistics, not only from an operational point of view, but also from a conceptual point of view. While it is true that there are international definitions, it is not always easy to measure the strict concept of a migrant. Moreover, there are increasingly different forms of migration, and there are is a variety of sources that provide partial measurement of migration.

79. The challenge for statistical offices is to combine data from the different sources into reliable and integrated information on migration. The 13 case studies showed the different ways and data sources that countries use to improve their migration statistics.

80. The Task Force conducted a survey to assess the potential of many different sources for producing migration statistics. According to the replies, most countries (31 of the 56

responses) use *more than one source for generating migration statistics*. Approaches differ because the circumstances and data sources vary greatly across countries. The overall tendency to involve more administrative sources in measuring migration applies both to countries that base their population statistics on classical methods (surveys, exhaustive censuses) as well as to countries that base their statistics on population registers.

81. The choice of sources for migration data often depends on *geography*. Countries that are relatively isolated or with fewer points of entry and exit (e.g. New Zealand, Malta) can produce migration statistics based on very different sources than others with many access points or without checkpoints, like countries in the Schengen Area in Europe.

82. Another important limitation when studying migration statistics is that there are *different definitions* of the migrant population in different administrative sources. In many cases, the standard time frame of 12 months of actual or planned stay is not the one used for the preparation of migration statistics, but other concepts (e.g. other time periods) are adopted. In addition, different administrative sources within the same country can offer information based on different concepts.

83. In the European Union countries, migration statistics are intended to be *fully consistent with changes in migrant stock*. In many other cases, migration statistics do not pursue this full consistency between flows and stocks. Moreover, many countries produce separate data on long-term migrants, asylum seekers, non-permanent foreign population, short-term migrants, etc.

84. In terms of data integration, there are some relevant examples of combinations of data *both at macro and micro levels*. As regards to micro-integration, that is, a combination of sources at the individual record level, there are several examples that can illustrate the trend for the future. The combining of records may generate overlaps and these are addressed through adjustment and calibrations processes. This integration is taking place to improve the measurement of stocks and also to produce more complete statistics of the flows.

85. The main weakness of the population registers is their *difficulty in detecting emigration*. If a person does not inform the authorities of their exit from the country, the administrative register has to be updated by administrative procedures that are not always easy or immediate, so the population registers may contain persons no longer residing in the country.

86. By combining sources, *'presence signals'* can be detected. Thus, a population register can be linked with tax or social security files to be more certain about the presence of people in the country. It is of particular importance in the combination of different administrative records in cases such as Latvia or Austria where logistic regressions are used for estimating resident population from evidence in different administrative registers. The analysis of the 'presence signals' may also help identifying persons who immigrated but are not in the population register (in Italy), and deceased persons (in Austria and Spain).

87. As for statistics on migration flows, there are some examples of integration in the strict sense, i.e., the combination of sources to obtain a single database. For example, in Hungary and Canada different sources are micro-linked. This micro-integration is carried out by linking information on foreigners across different administrative sources and with the population register, to avoid overlapping and to provide additional variables.

88. An initiative that is beginning to be demanded from users and would be very promising is *the exchange of individual level data* between countries for statistical purposes. The main implication of measuring migration flows is that they affect one country of origin and another of destination. So it is logical to think that the sharing of the statistical data will improve these statistics. The exchange of individual data for statistical purposes is likely to develop further, although it is necessary to overcome barriers, not just legal nor technical. Nordic countries provide a good example of such exchange.

VII. Recommendations

89. Based on the work carried out by the Task Force and the content of this document, a number of recommendations to national statistical offices are presented in this section. These recommendations are considered relevant to all countries, regardless of their current situation with the availability and use of administrative data for migration statistics.

A. Improve access to administrative data for national statistical offices

90. In countries where administrative records are not or only at some extent available to the statistical offices to produce official statistics, a priority is to take measures to facilitate the access to administrative records and generalizing their use by the statistical office. This may require support at the political level and, in some cases, changes in the legislation.

B. Use administrative data for migration statistics

91. Many administrative sources contain information that can be used in migration statistics. These include registers and databases on population, visas, passports, border control, residence and work permits, and asylum seekers and refugees. It is recommended to examine the potential of all these administrative sources for migration statistics and use them for the purpose of migration statistics to the extent they are fit for this purpose. This requires cooperation with relevant ministries, agencies and administrations responsible for the data.

92. It is advisable to produce several statistical outputs that provide information on migration stocks, as well as inflows and outflows, covering different types of migration flows. In some cases different statistics may provide not fully consistent data. However, their publication can help to give a global picture and to increase awareness that a single product does not provide all the necessary information on international migration. The simple publication of data from various sources is a first mechanism for integrating information on migrants.

C. Combine data from different sources using 'presence signals'

93. In those countries where administrative data are available for statistics, the combination of sources using 'presence signals' methods is very promising. The point in common to those different approaches is the usage of several sources to estimate a probability of an individual's residence in the country based on their presence in these sources. Combining data from different sources in such way allows to improve estimates of current population and stock of foreigners.

D. Pay attention to the quality of integrated data

94. As countries advance and standardize registers of social and population statistical datasets, data integration is more likely to take place at the unit record level across multiple sources. A greater attention to a quality assessment of the integrated data will necessarily become an integral part of the production of international migration statistics.

E. Be transparent about data integration methods used and develop standards

95. It is increasingly important for statistical offices to provide a transparent presentation of the rules and methods used for the data integration processes that produce the statistics of international migration flows and other types of international migration statistics. This opens up opportunities to develop methodological standards of data integration practices across the national statistical offices in the future. Individual countries will always have their own collection and data integration practice but there are likely to be more open technical forums for discussion of rules and methods used in the combining of sources for the production of migration statistics.

F. Promote international comparison and exchange of migration data

96. Another source of improvement can be found in the international comparison and in the exchange of data on migration, foreign born population and foreigners. Administrative sources typically underestimate emigration due to the low motivation of emigrants to register the event of moving out of the country. However, immigration data of countries where the emigrants go could cover these events. This can be checked using the so-called 'mirror statistics' that show the migration flow from country A to country B through the lenses of emigration statistics of country A and immigration statistics of country B. Furthermore, data from destination countries can provide information on the number and characteristics of emigrants and citizens of a country of origin. Exchanging data with other countries can thus be very useful for understanding migration processes and for improving their systems of measuring migration.

97. International data sources are very helpful for this. In the European case, Eurostat publishes quite detailed data of inflows and outflows that allow this comparison between countries of the European Union. In its Clearing House (database) of Migration Statistics, UNECE publishes such data from countries of Eastern Europe, Caucasus and Central Asia.

98. A breakthrough could be achieved if the databases were shared between countries to link at the individual level. There are already real cases like those of the Nordic countries. Today it is unlikely that exchanges of individual level data could be generalized, but there might be some initiatives between countries due to the intensity of their exchanges or others circumstances. The success of these exchanges could encourage others to follow the same path.

VIII. Further work

99. At the international level, further work could address the potential of using big data for migration statistics and the ways how the analysis of longitudinal data can be incorporated into annual migration statistics and complement the available cross-sectional measurements.

A. The potential of big data

100. The new and non-conventional data sources, such as mobile phones, credit cards and social networks — generally known as big data — could potentially be useful for migration statistics. Currently, statistical offices have great difficulty in accessing these data, which are in most cases privately owned. In the coming years, steps should be taken to gain access to these sources and find ways to use them for the purpose of official statistics. This information alone would probably be insufficient and biased, but when integrated with conventional sources it could improve migration estimates. It would be important to share

the emerging practices internationally, to support countries' first steps towards using the potential of such data for migration statistics.

B. Longitudinal data

101. Probably the future of data integration is not only in the improvement of quality in the measurement of migratory flows. The integration of data may allow new data on the migrant population to be available. The possibility of connecting population databases (or surveys) with files for employment, social security, taxes, health, education systems, among many others, will be able to provide more and better information in the future on living conditions or the social integration of the migrant population.

102. A clear example is found in longitudinal studies information. Official statistics focus almost exclusively on providing cross-sectional information on migrants. For example, there is barely any information about the duration of the migratory episodes (does a migrant live in the host country for a long time?) Or about very relevant time intervals when studying migratory flows: the time elapsed since arrival in the country of destination until obtaining a job, or between losing a job and emigrating. Longitudinal data are needed for addressing the question of whether migrant socio-economic outcomes improve or worsen over time. They are also needed to understand family migration and emerging patterns such as circular migration.

103. Sharing of good practices in creating datasets from the longitudinal perspective and using them in regular migration statistics would support NSIs in their work. Developing recommendations on how to harmonize such statistics internationally could be the subsequent step.

References

ESSnet (2008): Project on Integration of Survey and Administrative Data. Material available at: https://ec.europa.eu/eurostat/cros/content/isad-0_en

ESSnet (2011): Project on Data Integration. Material available at: https://ec.europa.eu/eurostat/cros/content/data-integration_en

ESSnet (2014): Project on Handbook on Methodology of Modern Business Statistics ('Memobust handbook'). Module on Macro Integration. Material available at: https://ec.europa.eu/eurostat/cros/content/macro-integration_en

Eurostat (2009): "Insights on Data Integration Methodologies". Proceedings of the ESSnet-ISAD workshop, Vienna, 29–30 May 2008. Eurostat Methodologies and Working Papers. Available at: http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-09-005

Eurostat (2013): "Statistical matching: a model based approach for data integration". Eurostat Methodologies and Working Papers. Available at: http://ec.europa.eu/eurostat/en/web/products-statistical-working-papers/-/KS-RA-13-020

SDMX (2009): "SDMX Content-Oriented Guidelines – Annex 4: Metadata Common Vocabulary". Available at http://www.sdmx.org.

UNECE High-Level Group for the Modernisation of Official Statistics (2017): "In-depth review of data integration". Document ECE/CES/2017/8 for the Conference of European Statisticians meeting of 19–21 June 2017.

United Nations ECE/CES/2018/3

Economic and Social Council

Distr.: General
9 April 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians

Sixty-sixth plenary session
Geneva, 18–20 June 2018
Item 4 (c) of the provisional agenda
Measuring international labour mobility

Measuring international labour mobility

Note by the Task Force on measuring labour mobility

Summary

This document presents an extract of the report *Measuring international labour mobility* prepared by a task force with members representing Israel (Chair), Germany, Italy, Mexico, Norway, Turkey, International Labour Organization, International Organization for Migration, Organization for Economic Co-operation and Development, United Nations Economic Commission for Europe, United Nations Statistics Division, World Trade Organization, and Goldsmiths University of London.

The current extract is prepared for translation purposes. It includes the report's introduction, the section on concepts of international labour mobility, overview of data availability, summary of case studies, conclusions and recommendations, and a description of issues for future work. The full version of the report, including four country case studies and a description of concepts of international migration and labour, is sent for electronic consultation among all members of the Conference of European Statisticians and is available at http://www.unece.org/index.php?id=47411.

GE.18-05554(E)

1805554

Please recycle

Contents

Chapter		*Paragraphs*	*Page*
I.	Introduction and background	1–9	3
II.	Concepts	10–53	4
	A. Stock of international migrant workers	17–21	6
	B. Non-resident foreign workers (residents working abroad)	22–47	6
	1. Overview	22–29	6
	2. Taxonomy	30–39	8
	3. Linking the taxonomy to existing statistical frameworks	40–45	11
	4. Data sources	46–47	11
	C. Flow of migrants including for-work migrants	48–53	12
III.	Overview of data availability	54–78	13
	A. Sources	54–55	13
	B. ILO methodological questionnaires on international labour migration statistics	56–57	14
	C. Availability of data sources on international labour migration and mobility from ILO survey	58–78	15
	1. Population census and household surveys	59–65	15
	2. Population-based registers	66–67	16
	3. Establishment surveys	68–69	17
	4. Employment services	70	17
	5. Border/admission statistics	71–73	17
	6. Register of nationals issued permits for work abroad	74	18
	7. ILO module on migration	75–78	18
IV.	Country case studies	79–85	19
	A. Structure	79–82	19
	B. Summary of data sources	83–85	19
V.	Conclusions and recommendations	86–100	20
	A. Conclusions	86–90	20
	B. Recommendations	91–100	21
	1. Main groups for measurement	91	21
	2. Variables	92	21
	3. Adjustment of surveys	93–94	23
	4. Integration of data from different sources	95–96	24
	5. Data exchange	97–99	24
	C. Future work	100	24
References			

I. Introduction and background

1. Globalization expands and accelerates the movement of information, capital, goods, services and people across international borders. This applies in particular to common market areas such as the European Union with free movement of goods, capital, services and labour. As it has become easier to travel and take up work in other countries, international labour mobility has greatly increased and has become a topic of growing policy importance. Policymakers have an interest in understanding both the short-term and long-term effects of labour mobility. However, while there is an increased demand for statistics on labour mobility, producing comprehensive and comparable statistics on the topic remains challenging.

2. Because the measurement of international labour mobility touches upon economic, social and demographic domains, labour mobility is considered from different viewpoints, which separately cannot provide a full picture. For example, migration statistics cover migrants but not those who work in another country without changing place of residence. They also more often measure long-term migration than short-term migration, and do not always include the purpose of the move. Tourism statistics measure short-term moves only. Labour market statistics often lack information on mobility. The improvement of statistics on labour mobility thus requires coordination across several statistical domains and data collection systems, at both national and international levels.

3. Recognising the importance of improving statistics on international labour mobility, the Bureau of the Conference of European Statisticians (CES)[1] reviewed this area in-depth in February 2015. The in-depth review identified three principal challenges facing producers of statistics on labour mobility: improving data comparability and coordination of work; enhancing accessibility of existing information; and filling data gaps (Statistics Austria, 2017).

4. Addressing the challenge of improving data comparability and coordination of work is necessary for several reasons. First, coordination between producers of labour, migration and tourism statistics within a country could help compilers of data on labour mobility to take advantage of the unique information produced in each domain. In addition, international harmonization of definitions and methodology for measuring labour mobility is important for assessing its overall global impact. Harmonization of concepts and improved data availability are required to answer to policy needs and to understand the changing economic realities surrounding labour mobility.

5. The second challenge of enhancing the accessibility of existing information should also be addressed. International organizations have taken important steps in aggregating data from different countries and making international comparisons. This challenge will also be met by countries themselves through better utilization of existing data sources and improved dissemination of these data.

6. Lastly, filling data gaps is another essential activity to be taken on by statisticians. These gaps can be filled through improvements in the collection of data through surveys,

[1] The Conference of European Statisticians is composed of national statistical organizations in the UNECE region (for UNECE member countries, see http://www.unece.org/oes/nutshell/member_states_representatives.html) and includes in addition Australia, Brazil, Chile, China, Colombia, Japan, Mexico, Mongolia, New Zealand and Republic of Korea. The major international organizations active in statistics in the UNECE region also participate in the work, such as the Statistical Office of the European Commission (Eurostat), the Organization for Economic Cooperation and Development (OECD), the Interstate Statistical Committee of the Commonwealth of the Independent States (CIS-STAT), the International Monetary Fund (IMF) and the World Bank.

administrative sources and improved data integration, though each come with unique challenges – including, but not restricted to, budgetary limitations. Data collection should be tailored to the specific requirements of measuring labour mobility as will be discussed in this guidance.

7. Based on the in-depth review, the CES Bureau established the Task Force on Measuring Labour Mobility in October 2015. The Task Force worked in conjunction with the concurrent International Labour Organization (ILO) Working Group on Labour Migration Statistics to minimize any possible contradictions and increase synergy between the two groups. While the ILO Working Group focuses on the concepts and definitions, the UNECE Task Force focused on countries' practices in the compilation of labour mobility statistics and the guidance that could be provided on this basis.

8. Members of the ILO Working Group and ILO staff have contributed to the present report. The ILO Working Group will prepare guidelines to be presented for adoption to the 20th International Conference of Labour Statisticians. These are expected to elaborate in greater detail the concepts and definitions related to labour migration and labour mobility.

9. The current document contains an extract of the report. It includes the report's introduction, the section on concepts of international labour mobility, overview of data availability, summary of case studies, conclusions and recommendations, and a description of issues for future work. The full version of the report, including four country case studies and a description of concepts of international migration and labour, is sent for electronic consultation among all members of the Conference of European Statisticians and is available on the UNECE web site at the following address: http://www.unece.org/index.php?id=47411.

II. Concepts[2]

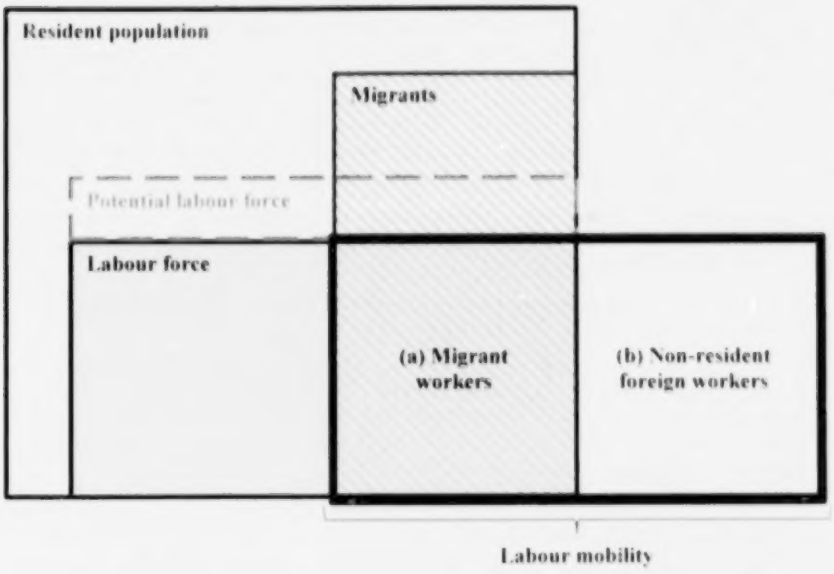
10. International labour mobility comprises all movements of natural persons from one country to another for the purpose of employment or supply of services.

11. Statistics on international labour mobility should rely on statistical frameworks in the areas closest to it, namely, international migration and labour. For information on the relevant concepts in these areas, the reader is referred to the full report that relies on the United Nations Recommendations on statistics of international migration (United Nations, 1998), the UNECE Handbook on the use of administrative sources and sample surveys to measure international migration (UNECE, 2016) and the ILO Resolution concerning statistics of work, employment and labour underutilization (International Labour Organization, 2013). The conceptual issues specific to international labour mobility are described below.

12. International labour mobility comprises all movements of natural persons from one country to another for the purpose of employment or supply of services.

[2] This section benefitted from the draft report "Towards improving statistical standards on international migrant workers" (International Labour Office, 2016) discussed at the 2nd meeting of the ILO Working Group on Labour Migration Statistics (Turin, 15–17 November 2016) and from the report on that meeting (International Labour Office, 2016).

Figure 1
Relation of the concepts of usually resident population, international migration, labour force and labour mobility



13. International labour mobility comprises all movements of natural persons from one country to another for the purpose of employment or supply of services.

14. Figure 1 provides a conceptual overview of the main population groups relevant to labour mobility and their relation with existing definitions of usual residence, migration and labour force. When looking at the population currently present in a country from the labour mobility perspective, the following two broad groups are of primary interest:

(a) "Migrant workers" who changed their place of usual residence from one country to another and are in the labour force of the destination country;

(b) "Non-resident foreign workers" who moved across borders for the purpose of employment or supply of services and who are not usual residents of the destination country.

15. The migrant workers may have arrived in the country for any reason. In measuring the in-flow of migrants, it is important to distinguish them by purpose of stay. This would allow identifying "for-work migrants", that is persons who arrived with a documented or declared intention of undertaking or seeking work at the time of entry into the destination country. While those who migrated for reasons other than work (e.g. as dependents, refugees or students) are not directly part of labour mobility at the time of their move, they may later contribute to the stock of migrant workers.

16. The following sections address the three main groups for the measurement of international labour mobility:

(a) Stock of international migrant workers;

(b) Non-resident foreign workers;

(c) Flow of migrants including for-work migrants.

A. Stock of international migrant workers

17. In order to define the stock of international migrant workers, it is necessary to put together the concepts of "international migrant" and "worker".

18. Regarding the aspect of work, the main issue is whether to apply the concept of "work" or "being in the labour force". Work is defined as "any activity performed by persons of any sex and age to produce goods or provide services for use by others or for own use" whereas labour force comprises the employed, that is, persons working for pay or profit, and the unemployed. It is pertinent to refer to the belonging to the labour force and define an "international migrant worker" as an "international migrant who is in the labour force of the country of which he/she is now a usual resident".

19. According to this definition, international migrant workers do not include people engaged in other forms of work than employment, such as own-use production, unpaid trainee work and volunteer work.

20. The migrant status and work status are defined in relation to the current moment. They are not related to the reasons for migration.

21. Some international migrants are not allowed to work for pay or profit. Others may be subject to restrictions limiting the type and/or location of the work they can undertake. At the same time, many of the international migrants who are subject to employment restrictions may want to work and in fact do so informally. On these grounds, the following two categories are included in the potential labour force:

(a) Those seeking but not available for employment because of restrictions related to their status as international migrant;

(b) Those available for but not seeking employment because of restrictions related to their status as international migrant.

B. Non-resident foreign workers (residents working abroad)

1. Overview

22. An important group of interest in labour mobility are those who engage in economic activity in the country concerned without changing their country of usual residence. From the perspective of the country where they work, they are "non-resident foreign workers"; from the perspective of their country of usual residence, they are "residents working abroad".

23. This group includes cross-border workers, non-resident service suppliers, seasonal workers, business travellers who receive remuneration from the country of origin, posted workers and others. Non-resident service suppliers, posted workers and business travellers are considered non-resident foreign workers if they stay in their country of destination for less than 12 months. Cross-border workers, seasonal workers and others are considered as non-resident foreign workers if they stay in the receiving country less than 3 months. If they stay at least three months, they are considered short-term migrants with usual residence in the country of destination.

24. Non-resident foreign workers could also include those who commute across borders to work without a formal permission to work in that country.

25. Information on both citizens and foreigners working under these arrangements are of interest when measuring labour mobility.

26. *Knowing* how many people are moving in the context of short term employment contracts or are providing services abroad (flows, stocks, as well as the number of associated trips) is very important for analysis of labour mobility. Given the interest that

labour mobility (migrant/non-migrant, employment/postings, short-term/long-term movements) generates amongst officials and research institutions involved in migration/labour/trade in services issues, these data would provide a useful gauge about the relative importance of such movements within wider economic international mobility.

27. An important criterion for distinguishing between different categories of people who arrive is whether the destination country considers the worker as an actual or potential usual resident. This is normally reflected in the type of legal residence status a worker is granted and often is implicit in the migration type. Seasonal workers, posted workers, cross-border workers, project workers and other such categories have limited rights and are not expected to change their usual place of residence. However, in some cases their status may change or lead to a continuous stay for more than a year, in which case they would be considered in the stock of international migrants.

28. As indicated in the in-depth review of labour mobility and globalization (Statistics Austria, 2017), in principle there is no need to create a new statistical framework, but rather build upon existing ones by defining a clear taxonomy of categories as outlined above and adjust for eventual gaps as necessary.

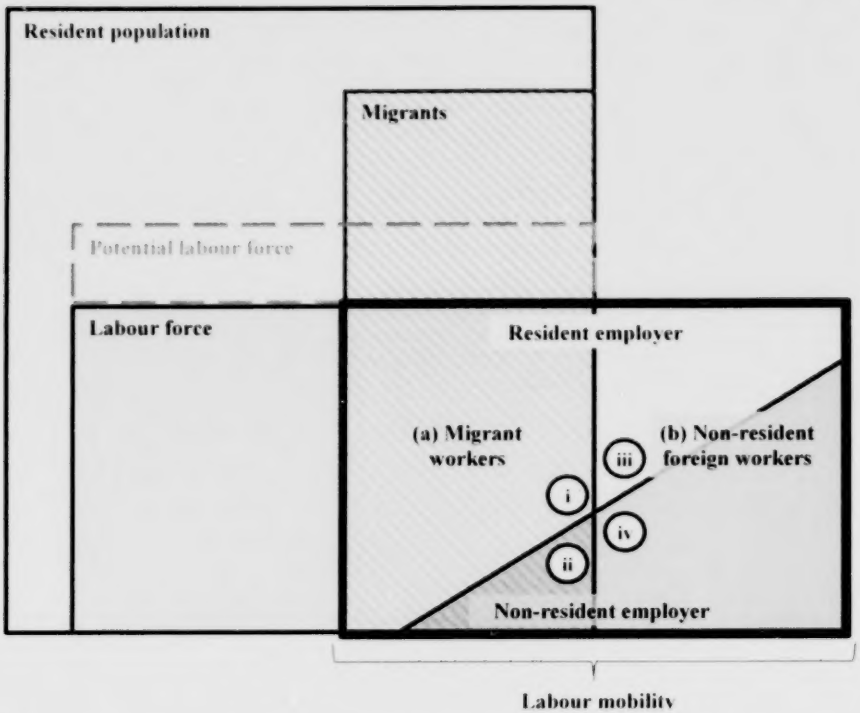
29. Statistical frameworks on migration and labour are more developed than those on labour mobility.[3] Concepts of service suppliers and people engaged in short-term employment in other countries, have only received attention in the context of statistical recommendations on trade in services.

[3] See work of ILO, OECD, Eurostat, IOM, UNECE, UNPD, UNDESA.

2. Taxonomy[4]

Figure 2

Groups of labour mobility in relation to the usually resident population and residence of employer



30. This section deals with non-resident foreign workers who are engaged in short term employment in the economy visited, regularly commute to another economy for employment (whether short or long term), or provide services based on a "service contract" and subsequently return to their home country. All these categories involve movement of persons across borders. Within the European Union (EU), the latter can commonly be referred to as "posting of workers" or "short-term assignments".[5] In the terminology of the General Agreement on Trade in Services ((GATS: General Agreement on Trade in Services, 1994) article I), paragraph 2, lit d) both these categories together are labelled as "mode 4" or "presence of natural persons". Harmonised statistics on this matter are scarce and it is therefore of utmost importance to address this phenomenon.

[4] This section is based on the in-depth review prepared by Statistics Austria for the Bureau of the Conference of European Statisticians (Statistics Austria, 2017), as well as chapter V of the Manual on Statistics of International Trade in Services 2010 (MSITS 2010) (United Nations, 2010) and chapter 16 of the MSITS 2010 Compiler's Guide (United Nations, 2014).

[5] The complexity of this phenomenon is well described by the fact that researchers have identified 12 more or less synonymous terms used in literature and they came up with a typology referring to 18 dimensions with more than 45 different categories of short-term assignments (Green, Baldauf, & Owen, 2009).

31. Figure 2 provides a conceptual overview of the scope of labour mobility and its relation with existing definitions of usual residence, migration and labour force. The triangle shape in the figure highlights labour mobility involving non-resident service suppliers. This cuts across the definition of usual residence established in population statistics as it relates to persons who are temporarily posted to a workplace abroad, such as consultants, temporary agency workers or self-employed service providers. Even if such an arrangement will typically be short-term, GATS does not imply any particular duration of stay. Although mode 4 implies international mobility of workers, the Annex on Movement of Natural Persons in GATS determines, "The Agreement shall not apply to measures affecting natural persons seeking access to the employment market of a Member, nor shall it apply to measures regarding citizenship, residence or employment on a permanent basis."

32. The residence of employers as defined in the Balance of Payments Manual 6th edition (BPM6) uses the criterion of one year. In that context, residence of institutional units is defined as follows:

"The residence of each institutional unit is the economic territory with which it has the strongest connection, expressed as its centre of predominant economic interest. [...] An institutional unit is resident in an economic territory when there exists, within the economic territory, some location, dwelling, place of production, or other premises on which or from which the unit engages and intends to continue engaging, either indefinitely or over a finite but long period of time, in economic activities and transactions on a significant scale. The location need not be fixed so long as it remains within the economic territory. Actual or intended location for one year or more is used as an operational definition; while the choice of one year as a specific period is somewhat arbitrary, it is adopted to avoid uncertainty and facilitate international consistency. [...] A household is resident in the economic territory in which household members maintain or intend to maintain a dwelling or succession of dwellings treated and used by members of the household as their principal dwelling. Being present for one year or more in a territory or intending to do so is sufficient to qualify as having a principal dwelling there. [...] As a general principle, an enterprise is resident in an economic territory when the enterprise is engaged in a significant amount of production of goods and or services from a location in the territory."

33. BPM6 Chapter 4 and MSITS 2010 Chapter 3 provide more information on residence of employers. In the present guidance, residence of foreign workers refers to usual residence (more than 12 months for long-term migrant workers and more than 3 and less than 12 for short-term migrant workers) whereas residence of institutional units/ employers refers to the concept of "predominant centre of economic interest" where one year is used as the criterion.

34. The combination of usual residence of persons with the residence of employers or place of employment yields four groups (i-iv in Figure 2):

(a) Migrant workers whose employment relation is with:

(i) A resident entity in the host country:

1. Long-term migrant workers, including self-employed persons: stay of at least 12 months;

2. Short-term migrant workers: stay between 3 to 12 months;

(ii) With a non-resident entity: stay of at least 12 months (e.g. long-term posting);

(b) Non-resident foreign workers whose employment relation is:

(iii) With a resident entity: stay of less than 3 months (e.g. seasonal or frontier workers);

(iv) With a non-resident entity or self-employed persons: stay of less than 12 months (e.g. short-term posting).

35. National accounts count the labour input of foreigners who work for resident producer units (groups i and iii in Figure 2) as contribution to the domestic value added. Balance of payment statistics distinguish the same groups by determining whether the work is based on an employment or a service contract. The latter is characterised by payments for a specific result without control over the method of work whereas employment relations imply remuneration for working time including control over the method of work. The Guide on the Impact of Globalization on National Accounts ((UNECE, 2011), paragraph 10.32) considers the definitions in the 2008 System of National Accounts (SNA) and the Balance of Payments Manual Sixth Edition (BPM6) (International Monetary Fund, 2009) as consistent with population statistics.

36. The Manual on Statistics of International Trade in Services 2010 (MSITS 2010) ((United Nations, 2010), paragraph 2.79) integrates the United Nations Recommendations on statistics of international migration, SNA 2008 and BPM6 as well as the definition of business travellers in the International Recommendations for Tourism Statistics (IRTS) (United Nations, 2008). The integrated classification is presented in Section C.2 of the Manual.

37. The MSITS breaks mode 4 down into 3 main categories for compilation of data:

(a) Contractual service suppliers, whether employees of a foreign service supplier (ii and iv) or self-employed (iv) and (i.1);

(b) Intra corporate transferees who initially work for a non-resident service supplier who has established a commercial presence, and are sent to work in the commercial presence. For simplification purposes it is assumed in this guidance that for short-term transfers (i.e. less than 3 months) the employment relationship remains with the non-resident supplier (sub-category of iv), for longer-term transfers (i.e. 3 months or more) it is with the receiving entity (sub-categories of i.1 and i.2)[6];

(c) Service sellers/persons responsible for setting up a commercial presence (part of group iv).

38. The following categories are of primary interest:

(a) Short-term migrant workers (i.2)[7] and non-resident foreign workers (iii) whose employment relationship is with a resident entity;

(b) Non-resident foreign workers whose employment relation is with a non-resident entity or who are self-employed (iv).

39. As foreseen above, an important factor to distinguish between categories (a) and (b) is where the employment relationship stands, or how to determine if a person is self-employed, providing services on behalf of his employer in the destination country, or in an employer-employee relationship in the host country. The *Guide on the impact of globalization on national accounts* details how important this distinction is in terms of impact on national accounts and productivity measurement ((UNECE, 2011) box 10.1).

[6] MSITS 2010 also includes foreign employees directly recruited by foreign established companies although admitting that these may also be considered as migrants.

[7] This category is mentioned here to complement stays of less than 3 months for employment purposes. It is not covered by the definition of non-resident foreign workers, as according to United Nations Recommendations on Statistics of International Migration, it is assumed that their place of usual residence changes for the period they stay in the destination country.

3. Linking the taxonomy to existing statistical frameworks

40. As stated in the *Guide on the impact of globalization on national accounts*, "the definitions and recommendations in the various statistical frameworks seem sufficiently harmonized to ensure that the population and labour statistics collected according to UN/ILO frameworks may be used in the national accounts and balance of payments without adjustments." In other words, the United Nations Recommendations on statistics of international migration (United Nations, 1998) and the ILO Resolution concerning statistics of work, employment and labour underutilization (International Labour Organization, 2013) are sufficient to cover the needs identified.

41. Other statistical frameworks are of particular interest here: the International Recommendations for Tourism Statistics 2008 (IRTS) and the Tourism Satellite Accounts: Recommended Methodological Framework 2008 (TSA-RMF) (United Nations, 2008). This section presents how these frameworks could be used and extended in order to develop relevant additional indicators on labour mobility for non-resident foreign workers.

42. IRTS is a comprehensive methodological framework for the collection and compilation of tourism statistics. This conceptual framework defines tourism and refers to related concepts such as country of residence, place of usual residence and usual environment. It introduces the activity of visitors from the point of view of their expenditure and presents the standard classification of products and productive activities that need to be used to conduct a comparable analysis of the demand and supply related to tourism. An extension of IRTS is the tourism satellite account, by which tourism statistics are linked with the mainstream of macroeconomic analysis (see TSA-RMF).

43. IRTS identifies international visitors as non-residents travelling in the country of reference, or residents travelling outside of it, being on a tourism trip. They must not be in an employer-employee relationship with an entity resident in the country visited and receive compensation for the labour input provided. This framework proposes to breakdown trips according to their main purpose, i.e. personal (holidays, leisure and recreation, education and training, health and medical care, etc.) and business and professional purposes.

44. The flows of visitors whose main purpose of trip is business and professional includes the activities of the self-employed and employees, as long as they do not correspond to an employer-employee relationship with a resident producer in the country visited, as well as those of investors, businessmen, or any other type of professional purposes. Although IRTS recognizes that for some countries, it may be difficult to distinguish between business and professional trips, it suggests that for some others it may be sufficiently important to disaggregate this information further. We will outline below how this framework could be used to derive relevant information on non-resident foreign workers whose employment relation is with a non-resident enterprise or who are self-employed.

45. As indicated above, the IRTS category "business and professional purposes" of trip is a particularly interesting category to analyse. For this, the category can be disaggregated into non-resident foreign workers whose employment relation is with a non-resident enterprise (stay of less than 12 months, e.g. short-term posting) and self-employed. Those correspond to contractual service suppliers (of which self-employed and employees). However, statistics resulting from this framework will not cover cases where there is a change of usual residence from one country to another and employment-related movements of less than three months (employer-employee relationship with a producer resident in the host economy).

4. Data sources

46. The potential data sources for statistics on non-resident foreign workers are enterprise surveys, administrative sources, household and labour-force surveys, population

censuses and border or passenger surveys. There are no dedicated and comprehensive data sources for such movements and stays, which means that compilers should select and when possible combine the following sources:

(a) Administrative sources, such as population registers, migration records, work permits, social security and employment documents. They are more pertinent in countries of destination. Entry and exit cards could be used as a sample frame for border/passenger surveys;

(b) Border/passenger surveys should be considered for obtaining characteristics of those traveling for business, work or employment purposes, combined with counts of border crossings;

(c) Household/labour-force surveys could help to obtain information for outgoing persons, in particular the self-employed, and if relevant combining this information also with counts of outgoing persons/trips;

(d) Enterprise surveys:

(i) Trade in services surveys would probably be the best source to collect data on non-resident foreign workers whose employment relation is with a non-resident enterprise, i.e. alongside the data on the value of the contracts;

(ii) Foreign affiliates statistics (FATS) data sources could be useful on intra-firm movements or direct recruitments by foreign affiliates;

(iii) Other enterprise surveys, covering specific services sectors, or temporary employment agencies could also be specially tailored to capture information of interest;

(e) Business registers may contain information needed to identify potential self-employed service suppliers.

47. Partner country data may provide useful information given that it may be difficult or impossible to collect details on characteristics of incoming persons or trips through surveys of enterprises or households, or using administrative records for outgoing persons. As a consequence, partner country information could turn out to be essential and be used/compared using a clearing house for such data. This implies that the data collection and quality of the compiling country depends on partners' willingness and capacity to collect and provide quality data, as this normally involves coordination, cost and in the case of international negotiations also certain interests, a central brokering institution may be important. Compilers therefore need to be cautious in using partner country data (definitions, national laws/regulations in place may differ, etc.).

C. Flow of migrants including for-work migrants

48. The flow of international migrants is the number of migrants entering or leaving a country over the course of a specific period, normally taken as one year. In the context of international labour mobility, the interest is in the flow of for-work migrants and their share all international migrants.

49. In the case of out-flow of workers and in-flow of foreign migrant workers, the concept is not that of currently participating in the labour force, but that of the documented or stated objective or intention of such participation in the future. In the case of return flow of citizens, the reference is not to current participation, but to participation in the past, in another country.

50. "For-work migrants" enter a country with the documented or stated objective of seeking or undertaking employment. Seeking or undertaking employment does not have to

be the only reason for the entry, nor even the main reason – it suffices for it to be just one of the reasons if there are several reasons for the migration. To measure this group, the reasons for the move have to be recorded and tabulated in the way that allows to identify whether seeking or undertaking employment was one of the documented or stated reasons for the entry.

51. In most situations, better data can be collected in the destination country on the in-flow than in the country of origin on the out-flow. Furthermore, the data sources are geared towards identifying the in-flow foreigners and less so of the in-flow (return) of citizens.

52. The reason or motive for the migration (including the intention to undertake employment) may also remain undocumented and undeclared. This may result from several factors. Firstly, the omission may be intentional, such as in the case of a person entering on a tourist or study visa but then staying on to seek or undertake work. Secondly, the intention may be overshadowed by some more powerful and immediate reason for the migration, as in the case of a person entering the country as a refugee or asylum seeker; many such persons have to seek employment, but that can hardly be the stated reason for their move. Thirdly, there may be no opportunity to state the reason for entry, as in the case of undocumented or irregular migrants. Fourthly, the same reason as above would apply also to migrants whose entry is not subject to controls.

53. From the perspective of linking the information on the flow to the stock of international migrant workers, it is to be noted that for-work migrants may or may not actually undertake work or continue to work and migrants entering with other objectives may or may not actually remain outside the labour force.

III. Overview of data availability

A. Sources

54. This chapter is concerned with empirical information on the current situation concerning data availability on international labour mobility in different countries. It addresses issues concerning data availability on the three main measurements relevant to labour mobility: (i) stock of international migrants and migrant workers, (ii) non-resident foreign workers and (iii) in-flow of migrants including for-work migrants, as defined in section II of this guidance.

55. The measurement of labour mobility touches upon various statistical domains using a variety of sources. In order to identify good practices, ILO prepared a set of questionnaires requesting information on the availability of statistics on the migration of workers to and from each country, and the sources and methods used in compiling such statistics.[8] In this chapter, we examine and draw implications from a broad analysis of the results of ILO questionnaires concerning data availability on the three main measurements of interest in labour mobility identified above. For the present purpose, the most useful indicators from the ILO survey analysed here concern the type of information that is available in different countries from different national data sources.[9]

[8] In 2015 the ILO circulated a series of methodological questionnaires on International Labour Migration Statistics as part of the Working Group on Migrant Workers. In July 2016 the ILO shared the data collected in the questionnaires with the UNECE Task Force on Measuring Labour Mobility, and the research on the questionnaire was presented at the second meeting of the Working Group on the Labour Migration Statistics, Turin, Italy, 15–17 November 2016.

[9] ILO plans to publish more detailed results from the enquiry in a separate publication.

B. ILO methodological questionnaires on international labour migration statistics

56. ILO circulated seven types of questionnaires (see Table 1). The questionnaires were sent to the national statistical offices (NSOs) of all ILO member States, with the request that each relevant questionnaire be completed by the appropriate unit of the NSO and/or the ministry or agency responsible for a particular data source. Depending on the data sources available, the same questionnaire could require response by more than one government ministry or agency. Not all countries answered all questionnaires and some countries responded more than once to a particular type of questionnaire (e.g. providing information on more than one household survey). Of the 190 countries[10] originally contacted, 128 countries returned at least one questionnaire, giving a total of 480 questionnaires.[11]

Table 1
Numbers of responding countries and questionnaires received

	Total		*Europe and Central Asia*	
Questionnaire	*Types of questionnaires received*	*Number of responding countries*	*Types of questionnaires received*	*Number of responding countries*
Q1 Population Census	122	122	48	48
Q2 Household Surveys	120	86	68	42
Q3 Border/Admission Statistics	49	48	20	19
Q4 Establishment Census	18	18	4	4
Q5 Establishment Surveys	18	16	10	8
Q6 Administrative Sources	126	72	89	40
Q7 Estimation of International Migrant Workers	27	26	11	10
All questionnaires /Countries	**480**	**128**	**250**	**50**

57. As we can see in Table 1, most questionnaires were answered once by each country. Only Household Surveys (Q2) and Administrative Sources (Q6) have significant differences between the number of questionnaires received and number of responding countries. For the purpose of analysis, the country questionnaires were consolidated in the following data sources (S1-S10). Each source is unique, in that at most only one of it is present in a country, irrespective of the number of questionnaires types it is based on.

[10] The term "country" is used in a general sense to indicate a country, territory or other relevant administrative unit suitable for the ILO data collection.

[11] Of the 190 countries and territories approached, 67 per cent responded, accounting for 70 per cent of the world population. The proportion of countries responding reaches 90 per cent in Europe and Central Asia, but remains below 50 per cent in the Americas and Arab State regions. However, in terms of the proportion of the population covered, Americas have one of the highest figures (84 per cent, USA is included, though not Brazil), while it is the lowest for Arab States (30 per cent). Overall, the ILO country survey has achieved very respectable response rates. The achieved level of response is actually better than the above figures indicate, given that the subject matter of the enquiry is information on *international migrant workers*, to which the contribution of China is much smaller than indicated by the size of its total population.

(S1) Population census (same as Q1)

(S2) All household surveys, including Labour Force Survey (LFS) (consolidation of all questionnaires of type Q2 returned by the country into a single source, S2)

(S3) Population and social security registers, other administrative sources (consolidation of all questionnaires in a subset of Q6 into a single source)

(S4) Establishment census and surveys (consolidation if Q4 and also questionnaire(s) in Q5 returned by the country into a single source, S4)

(S5) Private and public employment services (consolidation of all questionnaires in a subset of Q6 into a single source, S5)

(S6) Registers of international migrants, migrant workers, work permits issued (consolidation of all questionnaires in a subset of Q6 into a single source, S6)

(S7) Border/admission statistical sources (consolidation of all questionnaires of type Q3 into a single source)

(S8) Register of nationals issued work permit for abroad (consolidation of all questionnaires in a subset of Q6 into a single source)

(S9) LFS (the LFS questionnaire returned using Q2)

(S10) Questionnaire on the Estimation of International Migrant Workers in the Country (consolidation of all questionnaires of type Q7 returned by the country)

C. Availability of data sources on international labour migration and mobility from ILO survey

Table 2
Number of countries reporting a source, as percentage of all responding countries in a region

	Any (no.)	S1	S2	S1 & or S2	S3	S6	S3 & or S6	S4	S5	S4 & or S5	S7	S8	S9	Other
Africa	31	100	65	100	13	19	32	32	13	45	45	3	29	100
Americas	15	73	60	100	27	7	33	7	0	7	33	7	40	100
Arab States	5	100	100	100	40	0	40	40	0	40	20	0	80	100
Asia & the Pacific	19	100	32	100	5	21	26	11	11	16	37	5	11	100
Europe & Central Asia	44	95	91	100	55	39	73	18	23	34	41	5	77	100
Other small territories	14	100	43	100	43	29	64	14	7	21	21	0	29	100
Percentage of respondents	100	95	67	100	32	25	49	20	13	30	38	4	46	20

58. A number of pertinent observations on the pattern emerging from Table 2 follow.

1. Population census and household surveys

59. Population census and household surveys are the most widely available sources of data on international labour migration. All responding countries report at least one of these two data sources, with 122 (95 per cent) reporting a population census with relevant information, and 86 (67 per cent) reporting at least one household survey, including 81 (63 per cent) reporting the availability of sources of both types. The main merits of census as a source are that (i) the census can provide information on small domains (small areas and

small populations) because of its large size; and (ii) its coverage of the resident population tends to be more complete than that in typical sample surveys. The main merits of sample surveys arise from (i) their higher frequency (hence timeliness), and (ii) greater flexibility and richness of content.

60. Among household surveys, the labour-force survey (LFS) is the most common source (at 43 per cent), followed by income, expenditure or living conditions surveys (26 per cent).[12] This scenario has important practical implications.

61. In almost all cases, population census and/or surveys obtain information on basic demographic characteristics (age, sex, marital status) along with individual's employment status. A clear majority cover 95 per cent or more of the country's population, though under one-half (47 per cent) report covering refugees and asylum seekers. Information on country of birth and/or on the country of citizenship along with information on the person's employment status is available in a clear majority (90 per cent) of the countries, making it possible to identify international migrant workers in the population. In addition, over half (55 per cent) the countries report information on whether the individual has ever lived abroad.

62. Nearly half (47 per cent) report that the source(s) cover refugees apart from other non-citizens.

63. Nearly 75 per cent of the countries report information on country of citizenship, and 85 per cent on country of birth. Information on employment status is available in practically all countries. Information on country of birth and/or on the country of citizenship is available along with information on the person's employment status in clear majority (90 per cent) of the countries. Among these, in 65 per cent all the three items of information are available together in the same source: such information makes it possible to identify international migrant workers in the population. A similar picture pertains in relation to information on the migrant's previous country of residence and on duration since last entry (in-migration).

64. A majority (60 per cent) obtain some information on former household members who have moved out to live abroad for intended or actual duration exceeding a certain specified threshold.

65. One-third of the countries (34 per cent) report some information on remittances sent and/or received by international migrants. Most of these record remittances received by migrants from abroad, and only half of these record remittances sent back.

2. Population-based registers

66. Population-based registers containing information concerning international labour migration – which include general population and social security registers and other similar administrative sources, as well as registers of international migrants and of migrant workers, and work permits issued – are the second important source of information. One-half of the responding countries report the availability of at least one such source. The type of information on international migrant workers contained is of similar scope and type to that contained in population censuses and surveys. However, the availability of this type of source is quite different across regions: three-quarter of countries in Europe and Central Asia but only one-quarter in Asia and the Pacific report population-based registers containing information on international labour migration. Registers of international

[12] The 2014 EU-LFS had an ad hoc module on "Labour market situation of migrants and their immediate descendants" with 25 countries participating
http://ec.europa.eu/eurostat/statistics-explained/index.php/EU_labour_force_survey_-_ad_hoc_modules

migrants and of migrant workers, and work permits issued (S6) are also most prevalent in Europe and Central Asia (39 per cent), but otherwise the pattern is complementary (opposite) to the above for general population and social security registers.

67. These sources containing information on international labour migration include general population and social security registers and other similar administrative sources, registers of international migrants and of migrant workers, and records of work permits issued. Information on employment status along with on basic demographic characteristics is available in one-in-five (19 per cent) countries. The remaining countries report the availability of basic demographic characteristics only, without information on individuals' employment status. Whether the person has ever lived abroad is recorded in 35 per cent of the countries, i.e. in nearly three-quarters of the countries reporting the presence of the data source. Refugees and asylum seekers are included in less than half of the countries reporting the presence of the data source.

3. Establishment surveys

68. Even though the two sources are very different in terms of statistical methodology as well as the content, generally countries tend to report only one or the other of the two, either establishment surveys or employment services. Only a few countries in Europe and Central Asia report having both types of sources.

69. Establishment census and surveys (S4) with relevant information concerning international labour migration are reported by one-in-five (25/128) of the responding countries. Generally, the scope and content of the information relevant for international labour migration in sources of this type tends to be limited. For instance, one-half of the available establishment census and surveys do not cover the informal sector, and a similar proportion applies to the coverage of major economic sectors; only one-in-five of the available census and surveys cover establishments of all sector and sizes. Similarly, while the coverage of employment services, where they exist, is generally national, fewer than one-in-four state that refugees and asylum seekers are covered. In either type of source, only a minority (one-in-four overall) include information on employment status and country or citizenship or birth – which are among the basic variables required for the identification of international migrant workers. Practically no information is recorded on the migrant's duration since arrival into the country. Similarly, practically no information is reported on remittances sent or received by migrant workers.

4. Employment services

70. Private and public employment services (S5) are reported less frequently (by 13 per cent of the responding countries). They are most common in Europe and Central Asia (23 per cent) but practically absent in Americas and Arab States. Employment status along with information on country of birth and/or country of citizenship – needed for identifying migrant workers – is reported by 6 per cent only (8/128) responding countries. A similar proportion report availability of information on former household members who have moved abroad. Very few contain information on whether the person has ever lived abroad, and none report information on country of previous residence.

5. Border/admission statistics

71. These sources (S7) cover information from the country's border/admission statistics obtained from immigration/emigration forms at the borders and other points of entry and exit from the country. In contrast to the sources listed above which, by and large, concern migration stocks, border/admission statistics (obtained from immigration/emigration forms at the borders and other points of entry and exit from the country) constitute a major source of information on migratory flows in both directions. Border statistics constitute a major source of information, being available in a little under 40 per cent of the countries responding to the ILO survey; furthermore, a third of all responding countries report that

such statistics are maintained on a continuous basis. Generally, they cover all ages and contain information on sex and age of the individuals.

72. It is highly relevant to note that over 80 per cent of the exiting border/admission records cover special categories of migrants, including irregular and transient migrants, refugees and asylum seekers. On the other hand, only one in four among those contain information on employment status along with information on country of citizenship or birth and last country of residence – items necessary for the identification of international migrant workers.

73. Over half of the countries with border/admission records also report the coverage of nationals going to live abroad. In half of the cases, information concerning reason for departure, intended duration of stay abroad, as well as on destination country is collected.

6. Register of nationals issued permits for work abroad

74. Register of nationals who have been issued permits for work abroad (S8) are reported by very few counties, only 5 among the 128 responding countries (4 per cent). The country of destination is reported by 3 per cent (4 cases). The implication is that *available information on both flows and stocks of workers migrating abroad is very limited in most countries*. Partial information is available from population censuses and household surveys: a majority (60 per cent) of these obtain some information on former household members who have moved out to live abroad for intended or actual duration exceeding a certain specified threshold; one-in-three also report some information on remittances sent and/or received by international migrants. However, such information on out-migrating workers is likely to be seriously underreported.

7. ILO module on migration

75. ILO has developed a labour migration module, which can be added to existing household surveys, in particular to LFSs in countries of origin and countries of destination. The module was piloted in 2006–2007 in four countries (Armenia, Ecuador, Egypt and Thailand) and then implemented as part of the LFS in Ukraine and Moldova in 2012. The module provides a rich source of data including social and economic characteristics of migrant workers. The survey covers: the scale, scope and pattern of distribution of labour migration; socio-demographic composition of migrant workers (including educational attainment and training before departure abroad); and frequency, duration, economic activities and working conditions of migratory trips.

76. The Labour Force Migration Survey (LFMS) was conducted in the Republic of Moldova in the last quarter of 2012 in order to assess the extent of labour migration out of the country and to describe the demographic and socio-economic characteristics of migrant workers. Administered as a module of the regularly conducted Labour Force Survey (LFS), the LFMS was developed and implemented by the National Bureau of Statistics of the Republic of Moldova.

77. Conducted in the fourth quarter of 2012 as a module of the regularly administered LFS, the LFMS had a target population of individuals aged from 15 to 64 years who had (i) either left the country to work or look for work abroad within 24 months preceding the date of the survey or (ii) who intended to leave the country to work or look for work within six months following the survey. The survey used three separate questionnaires to gather information on migrant workers: (i) a questionnaire to collect information on household members living abroad; (ii) a questionnaire on household members who had not been abroad in the previous 24 months and who intended to migrate within six months following the survey date; and (iii) a questionnaire on household members who had been abroad in the previous 24 months and had returned and were residing in Moldova at the time of the survey.

78. Similarly, in 2012 the State Statistics Service of Ukraine, in collaboration with ILO, conducted the Survey on Labour Migration (SLM) in order to estimate the size of labour migration out of Ukraine and analyse the demographic and socio-economic characteristics of migrant workers.

IV. Country case studies

A. Structure

79. Four countries volunteered to prepare country case studies for the Task Force: Israel, Italy, Mexico and Norway. The Task Force established a common framework for the case studies, aligned with the main measurements relevant to international labour mobility (section II):

(a) Stock of international migrant workers with externally displaced persons as a sub-category;

(b) Non-resident foreign workers and residents working abroad;

(c) In-flow of migrants including for-work migrants.

80. The goal of the case studies was to collect and present current practices on the collection of data and preparation of statistics in these areas. They describe the availability of data for these measurements in each country, characterize the most important data sources as well as present a brief analysis of the effect of labour mobility on the local economy and labour force.

81. Each country compiled data tables according to the template specified in the framework as far as data was available. The data is the basis for the discussion on availability and the analysis presented in the below chapter. Full data tables will be made available at the UNECE website.

82. The full version of the report contains four country case studies from Israel, Italy, Mexico, and Norway.

B. Summary of data sources

83. The four case studies describe widely different contexts of international labour mobility and its data sources. The sources include household surveys, border surveys, population censuses as well as administrative data from entrance permits, residence permits, population registers, social security registers and employment registers.

84. For measuring migrant workers, non-resident foreign workers and residents working abroad, Israel, Italy and Mexico rely on surveys (Table 3). In Norway, registers provide the main characteristics of all groups except residents working abroad, whereas LFS provides additional variables of all groups except non-resident foreign workers. To measure the in-flow of migrants including for-work migrants, all four countries mainly rely on administrative sources.

Table 3
Data sources in Israel, Italy, Mexico and Norway

Group	*Israel*	*Italy*	*Mexico*	*Norway*
Migrant workers	LFS	LFS	1. Intercensal survey; 2. Survey on Demographic Dynamics	Registers; LFS
Externally displaced persons	1: Population and Immigration Authority; 2: LFS	Residence permits	1. Survey on Demographic Dynamics; 2. Commission for Refugee Assistance	Registers; LFS
Non-resident foreign workers	Palestinian LFS	EU-LFS	Survey of Migration on the South Border	Registers
Residents working abroad	LFS	LFS; census	1. Intercensal survey; 2. Survey on Income & Expenditure; 3. Survey of Occupation and Employment	LFS; Swedish registers
In-flow of migrants including for-work migrants	Population and Immigration Authority	Residence permits; municipality registers	1. National Institute of Migration; 2. Survey on Demographic Dynamics	Registers; (LFS)

Notes: Cells representing administrative sources are highlighted. Numbers refer to a source of primary, secondary or tertiary importance, respectively.

85. The studies also reveal the potential of using data from other countries and territories. To measure non-resident foreign workers, Israel uses Palestinian LFS and Italy proposes to examine the possibility of using LFS from other EU countries. Norway uses Swedish register data for statistics on their residents working in Sweden. None of the countries provide for a complete coverage of non-resident foreign workers and residents working abroad.

V. Conclusions and recommendations

A. Conclusions

86. The diversity of data sources presents many options but also many limitations when comparing data between countries. From that perspective, the following points need to be considered:

(a) Length of stay – periods used in definitions should be taken into account when making comparisons e.g. the number of months of residence to be included in a group and/or the intended length of stay versus the actual length of stay;

(b) Regular and irregular migrant workers – use of registers and administrative data are a large part of data collection – particularly for coverage of recently arrived migrants. These data may not take into account irregular migrants who may have bypassed border crossings and/or work permits. Surveys and censuses may better cover this population although population definitions must be considered here too;

(c) National needs – countries have various needs and are under different regional agreements that affect the population that they are interested in monitoring – at borders and places of work. These differences may affect current and future data collection. Care must be taken to allow fulfilment of national and international reporting needs as well as those arising for information on migration, labour or trade in services;

(d) Age – countries have different upper and lower age limits regarding participation in the labour force.

87. Change in place of usual residence is key in defining migrants, including for-work migrants. Whether a change in place of usual residence occurs, depends on a change in location and the actual or intended length of stay. As the case studies show, countries collect and use the variables country of birth and/or country of citizenship in addition to time of entry to identify migrants and migrant workers. This is in line with the United Nations Recommendations and very useful for one-time and long-term migrants.

88. Additional data on entries and exits are needed to capture the dynamics of non-resident foreign workers and residents working abroad. Separate identification of recurring entries/exits is needed to collect information on frontier or seasonal workers. Information on type of entry permit and length of stay permitted can be useful for dealing with posted workers or non-resident service suppliers if they allow their separate identification.

89. Particular attention must be paid to the place of work and type of work. Migrants often participate in the labour force differently from the native-born population, for example, by having a different rate of participation and working in different industries and occupations. To estimate the effect of migrant workers on a country's labour force and economy, it is important to distinguish between employees and self-employed and between resident and non-resident employers. The former distinction is common in labour-force statistics. Data that would allow distinguishing resident and non-resident employers are rarely collected and even when they are, the information is not published in statistics.

90. Administrative data on border crossings and visas and permits is the main source for measuring the flows of migrants and migrant workers. This source may be limited for countries that belong to areas with freedom of movement, such as the EU.

B. Recommendations

1. Main groups for measurement

91. Data collection should focus on identifying and characterizing three main groups:

(a) Stock of international migrants and migrant workers in the country – identified by country of birth and country of citizenship;

(b) Non-resident foreign workers and residents working abroad, clearly identifying separate categories based on the location of the employer;

(c) In-flow of migrants including for-work migrants to the country.

2. Variables

92. All groups may include regular and irregular migrants. The recommended variables are presented in Table 4.

Table 4
Variables recommended for data collection

Variable	*Comments*
Variables that identify migrants and non-resident foreign workers	
Country of birth	Possible answers may be: (a) list of all countries, (b) regionally relevant countries and "other", (c) country collecting data yes/no.
Country of citizenship	"Country" should be the country at time of data collection and not at time of birth, to account for changes in borders. Further aggregation is possible if necessary e.g. identifying foreigners as foreign-born with foreign-born parents.
Date of entry	Main aggregation by length of stay: less than 3 months; at least 3 and less than 12 months; 12 months and more. Other aggregations may be needed for analysis in other frameworks. Therefore it is important to collect date rather than length of stay. Intended length of stay may also be used as an indication of migrant status.
Purpose of stay	Legal purpose of stay from administrative sources or an answer to a survey question on the purpose of move. This information would allow identify for-work migrants as well as externally displaced persons.
Country of previous residence	For migrants.
Country of usual residence	For non-resident foreign workers.
Variables that relate to the labour force	
Labour-force status	Employed/ unemployed/ outside the labour force
Country of work	For residents working abroad. Possible answers may be: (a) list of all countries, (b) regionally relevant countries and "other".
Frequency or periodicity of commute	For example, daily/ monthly/ seasonally
Status in employment	ICSE-93
Industry	ISIC Rev.4
Occupation	ISCO-08
Education level	ISCED 2011

Variable	*Comments*
Variables that relate to the quality of employment	
Hours usually worked per week	LFS
Hours actually worked per week	LFS. Hours worked in the reference week of the survey
Hourly earnings	Can be used for the calculation of the pay gap e.g. between migrant and non-migrant or resident and non-resident workers (Indicator 1c1, UNECE 2015)
Exposure to physical health risk factors	Can be used for the calculation of the percentage of employed persons exposed to physical health risk factors (Indicator 1a3, (UNECE, 2015))
Exposure to mental health risk factors	Can be used for the calculation of the percentage of employed persons exposed to mental health risk factors (Indicator 1a4, (UNECE, 2015))
Variables that relate to employers	
Resident or non-resident employer	
Type of service supplied	EBOPS 2010
Relation between the employer and the client	Relevant for service suppliers. Intra-firm trade or not

3. Adjustment of surveys

93. A targeted survey or module can be a very useful source in understanding the patterns of international labour mobility. By measuring length and purpose of stay, labour-force status and other characteristics, they provide information on the different categories of migrants and non-migrants, including retrospective information on residents' employment abroad.

94. While it may not be possible to implement targeted surveys and modules frequently, providing data regularly on migrants and non-resident foreign workers and residents working abroad could be achieved by making the following changes to the Labour Force Survey (LFS):

(a) Broadening the investigation to persons who do not belong to the survey population to non-resident foreign workers regardless of length of stay and of whether the entity employing them is resident or non-resident. It is acknowledged that some countries may not be able to implement this recommendation because the LFS data collection is not designed to cover rare and difficult-to-access populations;

(b) Adding questions for those who work abroad on their country of work and on whether the entity that employs them is resident or non-resident;

(c) Changes in the survey plan to cover workplaces that serve as irregular residences;

(d) Adjusting interviewing technique by using face-to-face interviews instead of telephone interviews with households with migrants as household heads;

(e) Modifying sampling methodology to ensure that a sufficient number of migrants will be observed.

4. Integration of data from different sources

95. Use of administrative registers can supplement existing information sources and increase the value of the surveys already carried out. Countries should examine their administrative sources to see if they can supplement or replace some questions in LFS. For example, it may be possible to use social security data to identify whether an employer is resident or non-resident. The possibility to integrate data from other household surveys and border or passenger surveys should also be examined.

96. The integration of different kinds of data sources is a crucial step towards developing an effective system for the monitoring of migration. Many administrative sources contain information that can be used for statistics on migration and labour mobility. These include registers and databases on population, businesses, visas, passports, border control, residence and work permits, and asylum seekers and refugees. It is recommended to examine the potential of all these administrative sources and use them to the extent they are fit for purpose. This requires cooperation with relevant ministries, agencies and administrations responsible for the data.

5. Data exchange

97. Exchange of data between countries is encouraged in order to obtain a complete picture of labour mobility, particularly on non-resident foreign workers and residents working abroad. Without partner country data, it may be difficult or impossible to collect details on characteristics of incoming persons or trips. This implies that the data collection and quality of the compiling country depends on partners' willingness and capacity to collect and provide quality data, as this normally involves coordination, cost and in the case of international negotiations also certain interests, a central brokering institution may be important.

98. Compilers need to be cautious in using partner country data as definitions, laws and regulations may differ.

99. For measuring international mobility within areas with freedom of movement that do not require permits, such as the EU, data exchange among countries would be a particularly important source. A breakthrough could be achieved if the databases were shared between countries with a potential to link records at the individual level.

C. Future work

100. The full report describes how international labour mobility statistics could be improved through future work in statistics on establishments, social security, employment services, tourism, and undocumented migrants.

References

GATS: General Agreement on Trade in Services. (1994, April 15). Marrakesh Agreement Establishing the World Trade Organization, Annex 1B, 1869 U.N.T.S. 183, 33 I.L.M. 1167 (1994).

International Labour Office. (2016). Towards improving statistical standards on international migrant workers. Geneva: ILO.

International Monetary Fund. (2009). Balance of Payments Manual (6 ed.). Washington, D.C.: International Monetary Fund.

Statistics Austria. (2017. February). Labour mobility. 2. Geneva: UNECE Working Paper Series on Statistics. Retrieved from http://www.unece.org/statistics/working-paper-series-on-statistics

UNECE. (2011). The impact of globalization on national accounts. New York and Geneva: United Nations.

UNECE. (2016). Handbook on the use of administrative sources and sample surveys to measure international migration in CIS countries. New York and Geneva: United Nations. Retrieved from http://www.unece.org/index.php?id=44114

United Nations. (1998). Recommendations on statistics of international migration (Revision 1 ed.). New York: United Nations. Retrieved from https://unstats.un.org/unsd/demographic-social/sconcerns/migration/#docs

United Nations. (2008). International Recommendations for Tourism Statistics. New York: United Nations.

United Nations. (2008). Tourism Satellite Account: Recommended Methodological Framework. New York: United Nations.

United Nations. (2010). Manual on Statistics of International Trade in Services (Revision 1 ed.). United Nations.

United Nations. (2014). MSITS 2010 Compiler's Guide. United Nations.

United Nations ECE/CES/2018/4

Economic and Social Council

Distr.: General
26 March 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians
Sixty-sixth plenary session
Geneva, 18–20 June 2018
Item 4 (d) of the provisional agenda
Use of registers and administrative data for population and housing censuses

Guidelines on the use of registers and administrative data for population and housing censuses

Note by the Task Force

Summary

The document presents an extract of the *Guidelines on the use of registers and administrative data for population and housing censuses*. The Guidelines were prepared by a Task Force composed of representatives of the following countries and international organizations: the Netherlands (Chair), Austria, Canada, Estonia, Germany, Ireland, Israel, Italy, New Zealand, Norway, Poland, Portugal, Republic of Korea, Slovenia, United Kingdom, United States, Eurostat, United Nations Food and Agriculture Organization (FAO), United Nations Population Fund and UNECE.

The current extract of the Guidelines is prepared for translation purposes. It includes selected parts of the Guidelines: (i) introduction; (ii) scope of the new UNECE guidelines and definitions of register-based and combined censuses; (iii) essential features of a population and housing census; and (iv) considerations when transitioning from a traditional census to a register-based or combined census.

The full text of the Guidelines has been sent to all members of the Conference of European Statisticians for electronic consultation. It is available at: http://www.unece.org/index.php?id=47411. Subject to a positive outcome of the consultation, the CES plenary session will be invited to endorse the Guidelines.

GE.18-04680(E)

1804680

Please recycle

I. Introduction

A. Background

1. Between 2012 and 2015 the UNECE Steering Group on Population and Housing Censuses coordinated the preparation of the *Conference of European Statisticians (CES) Recommendations for the 2020 Censuses of Population and Housing*. The Steering Group managed the work of nine topic-related Task Forces established to prepare initial drafts of the various chapters of the Recommendations. The CES subsequently adopted the Recommendations in June 2015, and these are available both in electronic format on the UNECE website[1] and in printed form in English, French and Russian.

2. In October 2015, the CES Bureau conducted an in-depth review of the diversification of population census methodologies and sources, based on a paper by Finland and Turkey (ECE/CES/BUR/2015/OCT/3) and a note by UNECE (ECE/CES/BUR/2015/OCT/3Add.1). As an outcome of the review, the Bureau supported the preparation of new guidelines on the use of registers for population and housing censuses, and requested the Secretariat to prepare new terms of reference for the Steering Group on Population and Housing Censuses and for a Task Force on Register-Based and Combined Censuses (Report of the Bureau meeting: ECE/CES/BUR/2015/OCT/21).

3. A draft of the proposed new Guidelines produced by the Task Force was presented and discussed at the meeting of UNECE Experts on Population and Housing Censuses in Geneva in October 2017 and consequently revised in the light of comments made by countries at the meeting. This publication presents the Guidelines as subsequently agreed by the CES.

4. Before the guidelines are presented, the following section summarises the census methods adopted by countries in the UNECE region and their evolution over time.

B. Census methods in the UNECE region and their evolution over time

5. There are many different ways to conduct a population and housing census. For the sake of simplicity, this document summarises only the three main categories of census methods: the 'traditional' census, the 'register-based' census, and the 'combined' census. However, a more detailed discussion of the various census methodologies is given in the CES Recommendations.

6. The traditional census is here intended to mean a census based on the direct count of all individuals and the collection of information on their characteristics through the completion of census questionnaires, either in paper form or electronically. The information is collected in the field across the whole country in a relatively short period of time, normally no more than two weeks. Questionnaires can be completed either directly by the households (with delivery and collection of paper forms undertaken by enumerators, the postal service or other methods, or online in the case of electronic questionnaires), or by the enumerators during an interview of the household.

7. The traditional census has a number of disadvantages. First of all, it is a very complex and expensive operation, mainly due to the need to employ a large temporary workforce for the field data collection (enumerators, supervisors and managers), and to print, distribute, and process a very large number of forms. Moreover, in most countries there are increasing difficulties in enumerating certain population groups, particularly those

[1] http://www.unece.org/publications/2020recomm.html

characterized by high mobility and multiple residences, and an increasing reluctance of the respondents to be enumerated, for various reasons. Finally, the traditional census is normally conducted only every 10 years (because of its cost and complexity) and the results often only become available after a relatively long time after data collection, while many users would like to have timelier and more frequently updated information.

8. Some countries have addressed some of the disadvantages of the traditional census either by using sampling (where most households complete only a short form with basic information, while a sample completes a more detailed long form, thereby reducing the total amount of information collected and processed), or by facilitating an online self-response option, which may result in field cost savings and improved quality but requires very careful planning and implementation. Another approach is to spread the fieldwork over time and adopt sampling, as it is done in the 'rolling census' approach developed in France.

9. A totally different approach from the traditional census is the register-based census that was developed by the Nordic countries in the 1970s. Denmark was the world's first country to conduct a fully register-based population and housing census in 1981. Under this approach there is no direct collection of data from the population, and the traditional enumeration is replaced by the use of administrative data held in various registers (population register, building/address register, social security register, etc.) through a matching process, normally making use of personal identification numbers. This approach permits the production of census data at a much reduced cost and with relatively limited manpower, once a good quality system of statistical registers has been established.

10. Since the 1990s, a number of other countries in Europe have developed innovative methods to conduct the census, combining the use of administrative data with a limited collection of data from a field enumeration of the population for specific variables. Under this approach, called a combined census, the field data collection can cover the whole population or just a sample. Often this approach is adopted in the transition from a traditional to a register-based census.

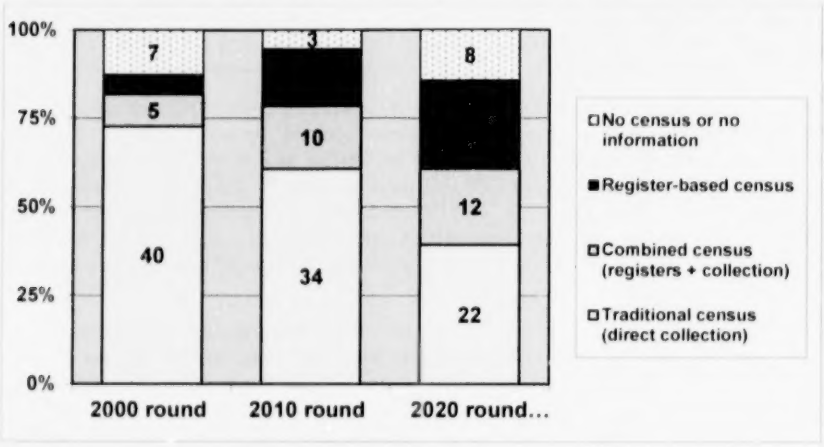
11. In the 2000 census round only few countries in the UNECE region[2] conducted a register-based or combined census (three and five countries respectively) and the traditional census was still by far the most popular approach in the region (40 countries)[3]. However, in the 2010 round, there was a significant increase in the number of countries conducting a register-based census (from three to nine) or a combined census (from five to ten), and a corresponding decrease of the traditional census (from 40 to 34 countries) (see Figure 1).

12. Based on information on tentative plans for the 2020 round, the trend of moving away from the traditional census continues: out of 48 UNECE countries for which information is available, 14 countries plan to conduct a register-based census (29 per cent), 12 countries are planning a combined census (25 per cent) and 22 countries are continuing with their traditional census (46 per cent). If only the 32 member countries of the EU (European Union) and EFTA (European Free Trade Association) are considered, then 13 countries plan a register-based census in the 2020 round (41 per cent), 9 countries a combined census (28 per cent), and just 10 countries will continue a traditional census (31 per cent).

[2] The UNECE region includes countries in Europe, North America, Central Asia, plus Turkey and Israel.

[3] Source: Valente, 2015, From the 2010 to the 2020 census round in the UNECE region – Plans by countries on census methodology and technology. Paper submitted to the Meeting of the UNECE-Eurostat Group of Experts on Population and Housing Censuses, Geneva, 30 September to 2 October 2015, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2015/mtg1/UNECE_paper_Paolo_draft_0925_rev2.pdf

Figure 1
Number of UNECE countries by census method in the 2000-2020 Census rounds



II. Scope of the new UNECE guidelines and definitions of register-based and combined censuses

13. The scope of these new UNECE Guidelines is not on traditional censuses, but on register-based and combined censuses. Therefore, only definitions of register-based and combined censuses are given. More information about traditional censuses can be found in the Recommendations for the 2020 census round. In these new UNECE Guidelines different kinds of registers (on persons and buildings) are noted with a focus on those used in censuses.

14. For some of the definitions we can refer to those presented in the UNECE publication *Register-based statistics in the Nordic countries*[4]. On paragraph 63 of that publication a ***register*** is defined as a systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up-to-date, correcting, or extending, the register, that is, keeping track of any changes in the data describing the units and their attributes. ***Administrative data*** sources are data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service. They include administrative registers (with a unique identifier) and possibly other administrative data without a unique identifier. ***Statistical registers*** are registers created for statistical purposes. They are typically created by transforming data from registers and/or other administrative data sources.

15. In some countries the term 'administrative data' is used as a synonym for 'register-based' data. In other countries a distinction is made between the two, and 'administrative data' is taken also to include administrative sources other than registers.

[4] http://www.unece.org/fileadmin/DAM/stats/publications/Register_based_statistics_in_Nordic_countries.pdf

16. A register-based census system is built around a set of basic registers that contain comprehensive data on the units that are to be described in the population and housing census (see para. 123 of *the CES Recommendations for the 2020 census round*). Some register-based census countries miss some of the census variables in all of the available registers and choose to support their census with unit record data (microdata) from an already existing sample survey. All register-based census countries have in common the fact that no specifically designed census questionnaires are used to collect information about the population. Therefore, register-based censuses are in general much cheaper than combined censuses and especially so compared to traditional censuses.

17. In a combined census, statistics are created by using registers and other administrative sources, together with information from either sample field data or full field enumeration for selected variables (see paras. 52 and 116 of the CES Recommendations for the 2020 census round).

18. The remainder of the Guidelines are organized as follows. Chapter III describes the essential features of a census and how these may be met by register-based or combined censuses. Chapter IV describes a number of elements that need to be taken into account when planning a transition from a traditional census to a register-based or combined census. The following chapters are available only in the full English version of this document. Chapter V outlines a common framework that describes the process of conducting these non-traditional censuses. Chapters VI - IX provide more details on the processes and methods associated with each aspect of the framework and the role that quality assurance plays at each stage. Chapter X, together with Annexes A-I, present case studies from a number of countries that have transitioned or plan to transition from a traditional full-enumeration census to a register-based or combined census.

III. Essential features of a population and housing census

19. The essential features of a population and housing census were originally defined by the International Conference of Statisticians as early as 1853 in Brussels, and by incorporating these features, countries have been able to carry out censuses that have been internationally comparable in terms of methodology and quality over time. Nowadays, these five essential features have been redefined and highlighted by the CES[5] with the aim of ensuring the coherence of census data gathered in different countries with different levels of technical development and different cultures. Adopting all these features – regardless of the methodology of data collection – enables NSIs (National Statistical Institutes) to collect population data of internationally comparable quality that allows making decisions and forecasts on the population development.

20. The five essential features of a census are:

(a) Individual enumeration;

(b) Simultaneity;

(c) Universality (within a precisely defined territory of a country);

(d) Small area data;

(e) Defined periodicity.

[5] Conference of European Statisticians Recommendations for the 2020 Censuses of Population and Housing (United Nations, 2015), see http://www.unece.org/publications/2020recomm.html, paragraphs 23-28.

How register-based and combined population and housing censuses can be designed to satisfy each of these features is discussed in the following sections.

A. Individual enumeration

21. The principle of individual enumeration is a fundamental feature for any census of population. Traditionally, this has been effected by providing questionnaires that ask questions of each individual within a household. In the case of register-based censuses a different approach is adopted where the data are taken from administrative registers. In such circumstances it is important that each census unit has a special, uniquely identified, record in the registers used. Then the registers become a useful source for the census. In the case of a combined census only some of the variables are derived from administrative data sources, and then this same approach is used for those variables.

22. If a single identifier for a particular unit does not exist across a range of registers, it is necessary to create a new statistical identifier (based on a group of identifying variables) to link the variables held in the respective registers, and to check carefully its quality (for errors and uniqueness).

23. Sometimes it is necessary to create the necessary census variable using information from several administrative registers and composing special algorithms for its calculation. This is possible if the units in all these registers are uniquely identified by the same identifier. In this case the variable created in such a way should be uniquely identified as well and saved in a statistical register. In the event that not all units are uniquely identified by the same identifier it may be possible to create a new statistical identifier as explained in the next paragraph.

24. The basic counting units of a population and housing census include not only persons, but also households, families, and dwellings (whether occupied or vacant). All of these units need identification, but there is no need to use five different identification variables. The minimal necessary identifiers are the one for persons (person ID) and the one for dwellings (dwelling ID). These IDs must be linked with each other (a dwelling ID is assigned to each person) and for each occupied dwelling the list of person IDs of people living in it must be given. The dwelling ID makes use of the address code, which may contain also spatial coordinates.

25. Information about households is usually collected on the basis of the housekeeping concept[6] by those countries conducting a traditional census. This definition can be achieved through asking questions on a survey or census, but is more challenging for countries conducting a register-based census. Many such countries instead use the household-dwelling concept, which considers all persons living in the same housing unit to be members of the same household. While adopting this definition has minimal impact on the total number of private households, it can have a larger impact for certain household types, such as one-person households. This bias in the number of private households and in the estimated structure of the household types depends on the traditions of the country and on living conditions. These challenges for register-based countries also extend to construction of families within households using relationship information.

26. In some countries (such as, Slovenia) a household register exists. The existence of such a register eases the organization of a register-based census, especially when household IDs are included in the register. Then there is accurate information available about which person ID belongs to which household ID. A household register might therefore improve the quality of a register-based or combined census significantly. However, the situation of

[6] See paras. 768-769 of http://www.unece.org/publications/2020recomm.html

Slovenia is an exceptional one. Ireland is researching the potential of using a decision tree algorithm to determine relations between people in the same dwelling so that the current housekeeping definition of the household can be continued. As part of the work to understand the impact of transitioning to its so-called Administrative Data Census, the Office for National Statistics (ONS) in the United Kingdom is currently exploring the potential impact on users of changing from a housekeeping concept to a household-dwelling concept.

27. Sometimes it is useful to use identification codes for other units such as enterprises and organizations. If these are linked with person IDs and dwelling IDs they form a helpful tool for deriving other statistics, such as on commuting between place of residence and place of work.

B. Simultaneity

28. The fixed census moment is the condition defining the simultaneity of the census data. Traditionally, to ensure this condition, the enumeration is carried out over a very short time frame, ideally during one day only. Though most modern day enumerations are conducted over a two-three week period, all the data collected should refer to a specified reference period. This essential feature should be respected also in the case of a register-based or combined census.

29. If the administrative registers in use are regularly updated, then it is necessary to fix the census period and to take the data from all registers with reference to this period. Sometimes the registers are updated regularly at some specific date such as the beginning of a year, and then it is possible to use this date as the census period, and the simultaneity of the census is guaranteed.

30. In the case of a combined census it is important that the census reference period mentioned in the questionnaires and the reference period of the information taken from the registers are the same or as close to each other as possible.

31. When several administrative registers are used in the census, it is important that all data taken from them have the same reference period. Usually, census variables derived via special algorithms take some time to calculate; hence those census variables are only ready for publishing sometime after the census period. For some specific variables in combined or register-based censuses different reference periods are defined for the particular administrative purpose of the register. Demographic data can normally be taken from population registers at the beginning of the year. However, labour force data might be more relevant somewhat earlier in the year - as fewer people tend to be in employment around Christmas and New Year's Eve. For some administrative registers, it may not be possible to have Census Day as a reference day. For example, education registers often have relevant education data (e.g. referenced to a day early in the academic year) that may differ from a chosen Census Day. In such cases, the NSI may wish to take education data with a reference day as close as possible to Census Day as a compromise.

C. Universality (within a precisely defined territory of a country)

32. To ensure the universality of the traditional census the questionnaires used in the enumeration process are the same for all households and individual persons. If there are questionnaires in different languages, it is important to check if their content and the meaning of all questions is exactly the same.

33. If the administrative registers used in the census are common for the whole country and all population groups, the condition may be regarded as being met. However, if there

are different administrative data in different areas or for different population groups (such as an urban population register and a rural population register or if different administrative data are held in different cities), then it is necessary to analyse the possible discrepancies between the different administrative sources and find a way to define common census variables using these different administrative sources. In this case plausible results can be derived from these newly defined variables (via an appropriate algorithm) in a statistical register.

D. Small area data

34. Providing a rich wealth of information for small geographic areas and small population sub-groups (generically referred to here as 'small area data') is a key objective for a census of any kind, as there is generally no other single source of comparable data.

35. In the context of register-based or combined censuses, small area census data can be derived from administrative data providing the coverage is high, preferably covering the whole population. If there are some small areas that are poorly covered, resulting in the administrative data lacking some information (and thus showing poor universality), it will be necessary to seek to improve the administrative dataset before it can be used as a source for the census. Such poor coverage is likely to be a problem for the every-day usage of the administrative data, and so it should be improved anyway. Improving the statistical register can sometimes be done by adding information from another source, providing that linkage through common IDs is possible.

36. Sometimes there might be special administrative data for some small areas (particularly for some small groups of people). Then it will be necessary to combine these different administrative sources (see section III.2.). If the result is satisfactory, this combination is useable. In the case of a combined census it is also possible to supplement the lack of information in administrative sources with a survey that, if necessary, may use different data collection methodologies, such as doorstep or telephone interviews, or self-completion with paper or online questionnaires, to suit different areas. However, where a sample survey is used, problems can still arise with the coverage of small area data. In such circumstances, users' needs regarding the required level of detail of the census outputs should be taken into account before any decision is made on the sample size of the census survey.

E. Defined periodicity

37. Nowadays, censuses are generally organized worldwide on a ten-year cycle. The United Nations recommend that countries conduct at least one census every ten years (between 2015 and 2024 for the 2020 census round). To meet European Union requirements, member countries were required (by an EU Regulation) to conduct a census in 2011, and will similarly have to conduct the next one in 2021. However, some countries (such as Australia, Canada, Ireland, New Zealand, and Slovenia) have shorter periods between censuses. The same underlying ten-yearly cycle should be kept in all censuses, regardless the methodology to provide international comparability. If a five-year period has been used, then, for EU member states, one of the two census years of the country should coincide with the census year fixed by Eurostat.

38. An advantage of a register-based census is, however, the opportunity to conduct the census more often than the usual ten-year cycle, as register data are permanently available and more regularly updated. It is advisable also to prepare census software in such a way that it is permanently ready for using for any reference date. Then the periodicity between censuses can be ten, five, or two years, or censuses could even be conducted annually.

Yearly updates of demographic data are a key objective in many European countries that now adopt different census methodologies.

39. It is also possible to produce some census updates with a shortened list of variables (but long enough to meet users' requirements). This could save resources. From this it follows also that in countries where a regular decennial census will be continued using a combined methodology, some updates with a shortened list of variables can more easily be done more often if the variables on the shortened list can be derived from administrative data sources.

F. Conclusions of the features defined by the Conference of European Statisticians

40. From the above it can be concluded that if a country has a good system of administrative registers that is consistent, easily useable and of high quality (that means, all units are uniquely identified using a common identifier), or if statistical registers can be built with equivalent quality from administrative data sources, then it is possible to organise a population and housing census that satisfies all required CES features.

41. If the list of administrative data sources cannot achieve the required quality for the whole range of census variables, it may still be possible and useful to derive census variables using a combined or register-based methodology with a shortened list of variables (including those covered in administrative registers) collected in between regular censuses.

IV. Considerations when transitioning from a traditional census to a register-based or combined census

42. The decision to move to a combined or register-based census is motivated because of the several advantages to be gained. However, the move needs to be carefully managed, and there are necessary conditions relating to data, technology, legal and stakeholder issues for a successful transition from a traditional to a combined or register-based census. This transition may therefore give rise to some challenges. These are set out in section IV.2 and can present significant obstacles for some countries.

A. Advantages

43. Conducting a combined or register-based census has a number of advantages and opportunities that are described below.

1. Lower per capita costs

44. Traditional censuses are very expensive. In many countries that conduct a traditional census it is common that the census costs are equivalent to about two annual budgets of the NSI. It is understandable, therefore, that governments put pressure on the institutes to cut these expenses, especially when other data sources are available.

45. If a combined census is conducted with full field enumeration for some of the variables, the cost savings made by shortening the census questionnaires are partly lost again in the effort required for combining the information from the field with administrative data sources. While the savings may be modest, the approach may still be preferable, particularly where the transformation process allows more data from administrative sources to be used in future censuses.

46. If a combined census is conducted without full field enumeration for selected variables much larger savings can be achieved. Practice shows that introducing a combined census could lead to reducing the costs by 22 per cent compared to a traditional census[7]. Moreover, if a register-based census is conducted no surveying is required specifically for the census and large savings can be expected. Several countries have proved that on average 98 per cent of the costs of a traditional census can be saved this way[8]. However, one should realise that such census cost savings can only be reached once the necessary access to the appropriate registers has been established.

47. Some countries have switched from a traditional census to a register-based census in one census cycle. If this is attempted, then all the costs of making the change must be met within the decennial period. More usually, however, such a move is done in several stages, often by first adopting an intermediate combined census approach before moving to a register-based census. The cost of the change can then be spread over two or three census cycles.

48. Moving to a combined approach and, especially, to a register-based approach contributes to a more cost-effective census. It is clear that meeting government-imposed budgetary constraints provides an incentive for such moves, even if registers are incomplete or of insufficient quality to be used as sources for the census. In such cases, however, it should be made clear from the beginning that a country should not attempt a move immediately and should continue conducting some form of traditional census. However, even when countries continue to do so, innovations making greater use of administrative data could help the NSI work more efficiently. It helps if the relevant public authorities make administration data sources available to the statistical institutes to produce proto-type register-based statistics. The government itself can help a great deal both by removing any legal barriers to data sharing and by subsidising the transition.

2. Quicker to conduct

49. In countries with an established register-based statistical system, the total production time to conduct a register-based census is much shorter than any other kind of census, due principally to the fact that no field enumeration has to be conducted. This, of course, does not hold for combined censuses in which some fieldwork is still necessary. Without fieldwork, the census may generally require less time in the planning stages (thus having the advantages of saving both time and money) while maintaining - or even improving - the time taken for the delivery of output. However, one has to realise that the first time a register-based census is conducted it may take more time than in later census rounds as census planning has to be set up anew.

3. Fewer problems with non-response and reduced response burden

50. With no need to conduct any field enumeration, fewer problems with non-response and a reduction in response burden to zero can be expected if only, or mostly, administrative data sources are used, and providing these are comprehensive and cover the whole population. With ever-increasing non-response rates being reported in international censuses and surveys this is going to be a more and more important aspect. Even if the move is to a combined census, the response burden on the population will also be lower, in particular if only sample field information is collected.

[7] Calculation based on PPP information in Table 7.2 of http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf

[8] Calculation based on PPP information in Table 7.2 of http://www.unece.org/fileadmin/DAM/stats/publications/2013/Measuring_population_and_housing_2010.pdf

4. Possibility of a continuous census

51. The more administrative data a country uses in its census, the better the possibilities of an annual or even more frequent census. In theory, with good quality administrative data updated outputs could be produced on a daily basis. Such real-time censuses may be something for the future, but more regular than decennial data are now becoming increasingly expected by users. As information from other sources can provide such real-time information, censuses will be expected to keep pace. As a by-product, it is easier for the NSI to keep the knowledge and IT-infrastructure up-to-date if annual census updates are produced.

5. Better cooperation between units within the NSI

52. In some NSIs the different divisions or directorates are often structured to perform in isolated silos without too many contacts with each other. If an NSI moves to a register-based statistical system there is potential for this silo structure to be abandoned. By moving from a survey-based statistical organisation towards a register-based system, the traditional one-to-one relationship between sources and statistics is replaced by an m to n inter-relationship across all statistical branches. A better cooperation of units within the NSI thus becomes essential. Moreover, by better integrating statistics, the coherence of the statistical framework within the NSI is improved.

6. More time and resources for innovations

53. Innovations are crucial for the long-term perspectives of an NSI. As introducing statistical registers in the process of producing official statistics saves both time and money, it becomes easier to innovate. The resources saved could, instead, be used to stimulate administrative and technological innovations so that the processes of data production remain up-to-date.

7. More flexible and responsive to new information requirements

54. If all data are stored appropriately, not only can the regular statistics be produced more frequently, but there is potential for the creation of new statistics to meet changing user needs. The NSI can then become more flexible and responsive to new information needs and increase its value to society. Although this may not be an aim in itself, it can lead to a greater level of user satisfaction.

B. Necessary conditions for a successful transition to a register-based or combined census

55. If a country wants to move to a combined or register-based census a number of conditions are necessary before information from administrative registers (and other sources) can successfully be integrated to create the underlying statistical register. A number of these necessary conditions can present some challenges to the NSI, and these are discussed in the following paragraphs.

1. Legal base

56. Whatever type of census an NSI conducts, it should be within an established legal framework. To be able to conduct a combined or register-based census, in particular, there must be legal provisions that prescribe the access to, and protection of, the administrative data. Such a legal base is normally enshrined in a Statistics or Census Act.

57. The NSI must have legal authority to access the relevant administrative data, ideally free of charge, from whatever public authority sources, preferably including personal

identifiers. To avoid legal uncertainty or dispute, it should be stipulated that the right of access applies except in legal cases pertaining to the protection of public order or the security of the country. A further key issue that should be addressed in legislation is the legal provision for the NSI to have some influence or authority in the creation, revision or deletion of those administrative data that are to be used in the statistical registers.

58. In turn, the NSI must have a legal obligation to protect the confidentiality of the administrative data it obtains, and to adhere to the 'one-way traffic' only principle, except under specific circumstances mentioned in the legislation. Indeed, the relevant legislation could do more by generally prohibiting other data controllers having access to data held on the NSI's statistical registers.

59. In certain countries, legal requirements may constrain how a census can be conducted[9]. In some countries, the NSI first started exploring administrative data sources and, thereafter, found a legal base to make register-based statistics possible. In other countries, the legal base was first established and, thereafter, register-based statistics were produced and published. To gain experience in moving progressively to a register-based census it is often simpler for NSIs to start with producing register-based statistics covering just a selection of those variables collected in a traditional census, though it should be noted that the legal obstacles to overcome may be no fewer.

60. It is always the case that moving to a census methodology where administrative data sources play a role needs careful preparation including, in particular, pilot studies. NSIs should realise that once (part of) the fieldwork operation for the census is abandoned, reinitiating it becomes rather difficult. After the passage of time, the knowledge of how to conduct a traditional census is lost, and especially so in the case of a register-based census where no fieldwork is undertaken at all.

61. Legislation on access to administrative data may need to be supported by policies and directives that are internal to the NSI and that translate legislative requirements and central government policies and directives into requirements and responsibilities for the managers and employees of the NSI.

2. Public approval

62. While the law might give a legal licence for the NSI to a combined or register-based census, public approval is also necessary to ensure that such a census is acceptable. This might be more difficult to achieve than establishing the legal base. While in some countries people may get the impression that 'big brother is watching you' in the course of a traditional census, in some other countries using and linking administrative data collected for non-statistical purposes may be seen to be even more intrusive as the public has no control at all over the information about them that is to be disclosed.

63. In a traditional census, privacy concerns may lead to lower response rates or the deliberate giving of wrong information. It is getting more and more difficult to correct for such unit and item non-response. So, on the one hand the public might prefer a situation where fewer questions are asked if the equivalent information is already available. On the other hand, part of the population may prefer to answer census questionnaires directly instead of having their information taken from, and combined with, several administrative sources.

64. In a register-based or combined census people may feel uneasy about, or even object to, information from different administrative data sources is reused and linked in a census. It may not be clear to them that in the census the information is only used for statistical purposes. If no more census forms have to be filled in and only registers are used for the

[9] Particularly where representation in the national legislature depends on census results.

census, the public will generally be less aware that a census is being conducted. However, the absence of any public reaction should not be misinterpreted as public approval.

65. It is desirable, therefore, to prepare for possible specific questions on privacy, confidentiality and security issues in conducting a register-based or combined census. In combined censuses a discussion can be expected about which variables are to be included on the census questionnaires and which variables are to be derived from administrative data sources.

3. Stakeholder approval

66. Stakeholders - or, more specifically, data users - typically want each census to provide at least the same level of detail of information as in the previous census. However, this is not always possible when the census methodology changes.

67. It is important to inform and consult stakeholders beforehand. Users in particular can become critical if their expectations are not met. Even well-informed users can become very critical if they believe that they are going to lose some of the information that they had access to in the previous census. However, it is usually not possible to satisfy all users, and disappointment among some of them is often unavoidable when adopting a new census methodology.

68. It is important, therefore, to have a communication strategy for stakeholder engagement that should encompass some, or all, of the following goals:

- Create a transparent environment concerning the plans of the NSI;
- Assure users that their requirements will be taken into consideration;
- Inform stakeholders of the benefits of using administrative data and demonstrate that the information will continue to be kept secure;
- Strengthen partnerships with the stakeholders so that the NSI can benefit from outside expertise;
- Make stakeholders part of a successful transition to a new census approach.

69. Openness and the clear identification of new opportunities and benefits for the stakeholders will help to gain their approval. It is particularly important that there should be adequate consultation on any change in the provision of those statistics that have financial consequences for stakeholders (such as transfers of money to municipalities).

4. Cooperation between the NSI and other authorities

70. Good cooperation between the NSI and other (mainly government) authorities is vital in using administrative data sources in the census. The NSI needs to know when microdata (the administrative unit records) and the accompanying metadata can be made available before any register-based statistics can be produced. In a combined census, and even more so in a register-based census, the NSI is heavily dependent on administrative data holders to comply with their agreed or legal obligation to provide good quality data on time. If data holders fail to deliver, it is usually the NSI that is held responsible for the failure to publish census statistics on time.

71. It is vital to inform administrative data holders how important their data are for the NSI and how their data are to be used. Additional to a legal base (see IV.2.1.) and good contacts with other authorities, signing cooperation contracts or service level agreements could help in supporting the census process. In theory, administrative data sources from non-governmental authorities could also be used in the census, but this often creates

privacy and data quality concerns and involves commercial considerations: private sector data are more often than not only acquired at a substantial cost to the NSI.

5. Comprehensive and reliable statistical register system

72. A statistical register system that is comprehensive and reliable (in that it contains accurate and timely data) is essential for conducting a combined or register-based census. Administrative data sources, including administrative registers such as a population register, are not normally set up for statistical purposes such as conducting a census. A transformation process is therefore necessary in order to create a reliable statistical register system.

73. To assure the use of register-based statistics in official statistics it is important to have good working relations with administrative data holders. Conditional on the provisions of the legal base, and if the administrative bodies are conducive to it, in some countries there is also a potential to improve the relations between the data holders and the NSI by introducing new, or extend existing, register-based statistics, such as longitudinal studies to evaluate policy implementation. Of course, there has to be contact between the NSI and the relevant administrative data holders whenever an administrative data source is introduced as a new source. However, permanent contacts, facilitated for example via account managers, are vital to keep both the administrative data holders aware of the important role that their data play and the NSI informed about any changes to the microdata and metadata they receive. Only with regular contact between the NSI and the administrative data holders can the success of register-based statistics be maintained.

74. In many register-based census countries the system of administrative data sources is used by many different public authorities. The more users this system has, the better the quality one can expect. In using such a system for the census it is the quality of the resulting statistical register rather than the quality of the underlying administrative data sources that counts: are the data of good enough quality on which to base reliable census outputs?

6. Unified identification system

75. A unified identification system across different administrative data sources greatly facilitates register-based censuses. It is preferable to have unique ID-numbers at the unit record level that are common across all registers. For countries where unique ID numbers for persons do not exist, the ability to link data efficiently and accurately is a particular challenge.

7. Knowledge of administrative sources

76. When a country wants to move from a traditional census towards a combined or register-based census, building up a wide-ranging knowledge of the data held in administrative sources is important before actually making the move. Although building up knowledge can be effected in stages well before the census planning, the effort needed to make this process successful should not be underestimated. Many lessons about failures and successes can be learnt from countries conducting combined or register-based censuses, but the national context should always be taken into consideration. It is never advisable for an NSI to simply adopt the methodology of another country when setting up a combined or register-based census. However, by learning from the experiences of others the transformation period can be shortened drastically.

8. Transparency

77. If there are planned moves to a different census methodology, it is good practice to be transparent and share with stakeholders information on plans and tests as much as possible. As discussed in section IV.2.3 above, it is particularly important to inform users of any decision to move towards a register-based census as such an important change in methodology is likely to have an impact on the content and availability of output. Transparency and openness facilitate external review and feed-back on the new processes.

C. Difficulties that may arise

78. Despite the advantages noted in section IV.1 above, conducting a combined or register-based census has a number of disadvantages and risks that are described below.

1. Dependency on public authorities

79. In moving to a combined or register-based census, the NSI becomes heavily dependent on the public authorities holding the administrative records being used. NSIs have to realise that, for such authorities, the production of statistics is not a core activity to which they would normally give priority. For the NSI, any failure or shortcomings in the administrative registers will affect the quality of the derived official statistics, for which it must take responsibility.

2. Differences in concepts and definitions

80. Registers and other administrative data sources often adopt different concepts and definitions of population-related variables than those that generally apply in traditional censuses. NSIs should be aware that such differences may exist and decide whether these differences are acceptable when moving from a traditional to a combined or register-based census. What may, in one country, be considered an acceptable difference when assessing the balance between the continuity and coherence of the resulting statistics and the reduction in field costs, may be considered unacceptable to users elsewhere. NSIs should weigh up the balance before deciding whether they are willing to pay this price when moving towards a register-based census or a combined census without full field enumeration for selected variables. Sometimes, original definitions and concepts can be approximated rather accurately by derivations from different sources or by editing information from newly acquired census sources. However, this is not always the case and the NSI should then weigh up the balance between the acceptability of the differences and the costs of continuing full field enumeration for selected variables.

3. Timeliness of administrative registers

81. Public authorities responsible for maintaining administrative registers do not hold the data for statistical purposes, and, as a result, will have other priorities that could cause delays in the delivery of the relevant administrative data and metadata to the NSI. This can cause issues for the NSI regarding the timeliness of their register-based statistics, particularly where the timeliness of the delivery of data from different sources varies considerably.

4. Different reference periods

82. A particular problem that NSIs encounter when moving towards a combined or register-based census is that different sources of administrative data often have different reference dates. Sometimes a source gives the option of distinguishing clearly between reference dates and dates of events, but this good practice does not always apply. If these

problems cannot be resolved sufficiently, the risk is that not all sources will be harmonised to the same reference date. Then the question 'What is an acceptable difference in reference dates?' arises. However, the answer to this is dependent on the variable concerned. Some variables are rather stable over time and then a small difference in reference date is normally not a problem. Large differences in reference dates are always unwanted. Finally, it is relevant to realise that also in the case of census questionnaires not all information may in practice refer to the single reference date of the census. Especially in case the census information is given on a moment further away from the reference date recall effects may play a role and respondents do not always give the answers specific to the census reference date.

5. Privacy and security concerns

83. Using administrative data for purposes other than those for which the information was originally obtained inevitably leads to privacy and security concerns. These concerns often relate to the linkage of personal data from different sources. In some countries the legal framework has been specifically adapted to provide for this, suggesting that there is public approval (or at least acceptance) in the use of administrative data in official statistics. For other countries such a consensus has not yet been achieved.

6. Difficulty in identifying sub-populations

84. In a census it is important to ensure universality. However, the range and detail of outputs in a register-based census will be limited to those variables that can be derived from existing sources. These may not all relate to the entire population. Moreover, even for those countries that use sample surveys to collect data on information not available in administrative sources, it is sometimes difficult, or even impossible, to produce accurate outputs for small areas or specific sub-groups of the population because of the size of the sample population.

7. Keeping knowledge and IT infrastructure up-to-date

85. In countries that conduct census projects with large gaps between them, it may be difficult to retain staff within the NSI with the necessary experience and expertise to keep the knowledge and IT infrastructure up-to-date during the inter-censal period. However, when yearly census updates are introduced this difficulty is minimised.

8. Diminishing interest

86. In countries where censuses are carried out using questionnaires, not only the users but the general public itself will be interested in knowing the results. However, in register-based census countries, where people no longer complete census forms, there is often a decline in public interest in census results. Many people will not even be aware that a census has been taken and, as a consequence, national interest in the census is greatly reduced. Users will still have an interest in the statistical outputs, though evidence from register-based census countries suggests that their interest in the choice of original sources and the methodology used to produce the census data diminishes over time. This is also due to the fact that in those countries other outputs are often much quicker available than census results.

V. Chapters available in the full version of the Guidelines

87. The following chapters are available only in the full English version of this document: Chapter V outlines a common framework that describes the process of conducting these non-traditional censuses. Chapters VI - IX provide more details on the

processes and methods associated with each aspect of the framework and the role that quality assurance plays at each stage. Chapter X, together with Annexes A-I, present case studies from a number of countries that have transitioned or plan to transition from a traditional full-enumeration census to a register-based or combined census

Annex.

Glossary of terms and definitions

Accessibility: A measure of data quality relating to the conditions and modalities by which users can obtain, use and interpret the data.

Accuracy: A measure of data quality relating to the closeness of estimates to the unknown 'true' values.

Activity register: A register that holds information about residents' different activities that indicate a presence in the country or area. Such activities can include information on, for example, employment or other economic status, receipt of benefits or pensions, or student status.

Administrative data: Data holdings that contain information collected primarily for administrative (not research or statistical) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service.

Anonymization: The process of protecting the confidentiality of personal information by removing all unique identifiers from the unit records.

Coherence: A measure of data quality relating to the degree to which the census data can be combined in different ways and for various purposes with statistical information from other sources.

Comparability: A measure of data quality relating to the degree to which statistics are comparable between geographic areas and over time.

Combined census: A census in which some information on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes, but where other information that is not available from such sources is collected directly from individual persons and households by means of full or partial field enumeration or from other sample surveys.

De facto census: A census based on a count of persons at where they were present on the reference date.

De jure census: A census based on a count of persons at their place of usual residence on the reference date.

Deterministic method: A method without a random component that thus always leads to the same outcome

DSE: Duel System Estimation - a statistical method, based on a capture-recapture technique, applied to estimate the size of a population.

Hypercube: A high-dimensional statistical tabulation of, typically, four or more dimensions

LAU: Local administrative unit. The classification of local administrative areas used by Eurostat. Levels 1 and 2 equate to those areas that were previously classified, respectively at the NUTS 4 and 5 levels.

Metadata: Information about the content, structure, quality and other relevant characteristics of a register

Microdata: As used in these Guidelines, information in a register relating to a single entity or entities

Numerical address: Code linking to the address

Population register: A register of residents of the country

Probabilistic method: A method with a random component that thus not always leads to the same outcome.

Process quality: The quality of a statistical process as evaluated by the methods used, cost effectiveness and response burden.

Punctuality: A measure of data quality relating to the delay between the date of the release of the results and the target date (the date by which the data should have been delivered).

Register: A systematic collection of unit-level data organized in such a way that updating is possible. Updating is the processing of identifiable information with the purpose of establishing, bringing up-to-date, correcting, or extending, the register, that is, keeping track of any changes in the data describing the units and their attributes.

Register-based census: A census in which the data on the numbers and characteristics of the population are derived from information taken from administrative data sources held for non-statistical purposes. No information is collected directly from individual persons or households.

Relevance: A measure of data quality relating to the degree to which statistics meet current and potential needs of the users.

Statistical register: A register created for statistical purposes. They are typically created by transforming data from registers or other administrative data sources.

Synchronisation: The transmission of data recorded on handheld devices in the field to a central server.

Timeliness: A measure of data quality relating to the period between the availability of the information and the event or phenomenon it describes.

Traditional census: A census based on the direct count of all individuals and the collection of information on their characteristics through the completion of either a self-completion or interview-based questionnaires, either in a paper or electronic format.

Unique ID/key: A single alpha numeric identifier that relates a characteristic or variable to a particular entity (person, household or dwelling) across a range of different registers or administrative data sources.

XML: Extensible markup language. In computing a language that that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

United Nations ECE/CES/2018/5

Economic and Social Council

Distr.: General
3 April 2018

Original: English

Economic Commission for Europe

Conference of European Statisticians
Sixty-sixth plenary session
Geneva, 18–20 June 2018
Item 4(f) of the provisional agenda
Entrepreneurship statistics

Guidelines on the use of statistical business registers for business demography and entrepreneurship statistics

Note by the Task Force on entrepreneurship statistics

Summary

The document provides an extract of the *Guidelines on the use of statistical business registers for business demography and entrepreneurship statistics*. The purpose of the Guidelines is to provide practical guidance on how to develop and maintain the statistical business register to support the production of business demography and entrepreneurship statistics.

The Guidelines were developed by the Task Force on entrepreneurship statistics, composed of representatives of the Netherlands (Chair), Austria, Canada, Estonia, Germany, Ireland, Israel, Italy, New Zealand, Norway, Poland, Portugal, Republic of Korea, Slovenia, United Kingdom, United States, Eurostat, United Nations Food and Agriculture Organization, United Nations Population Fund and United Nations Economic Commission for Europe.

The current extract of the Guidelines is prepared for translation purposes. It includes selected parts of the Guidelines: (i) introduction; (ii) background and acknowledgements; (iii) importance of measuring entrepreneurship; (iv) statistics for measuring entrepreneurship; (v) business demography and international comparability; (vi) aims of the Guidelines; and (vii) overview of the Guidelines.

The full text of the Guidelines has been sent to all members of the Conference for electronic consultation. It is available at: http://www.unece.org/index.php?id=47411. Subject to positive outcome of the consultation, the Guidelines will be submitted to the 2018 plenary session of the Conference of European Statisticians for endorsement.

GE.18-05155(E)

1805155

Please recycle

I. Introduction

1. Business demography and entrepreneurship statistics provide information about the creation, survival and dissolution of businesses and about the characteristics and activities of entrepreneurs. Over the previous decades, interest in these types of statistics has increased as they are sources of information about innovation, competitiveness, economic growth and job creation. Such information is used to inform business policies, for economic planning, and for analytical and research purposes.

2. The production of business demography and entrepreneurship statistics of suitable quality requires good data sources. The statistical business register (SBR), which plays a central role in the production of economic statistics, can provide the statistical infrastructure required to produce business demography and entrepreneurship statistics in the form of structured data on business units and their characteristics and activities. The SBR can be the sole source of data, or may be combined with data from other existing sources. Use of the SBR in addition to, or instead of, other sources, provides an opportunity to reduce costs, increase efficiency and improve the quality of the resulting statistics.

3. The Guidelines provide practical guidance on how to develop and maintain SBRs to support the production of business demography and entrepreneurship statistics. The Guidelines contain definitions and compilation guidance for key demographic events, and they propose business demography indicators that can be produced directly from the SBR. They also describe the requirements for the SBR to produce business demography indicators and they provide guidance on linking the SBR with other data sources.

4. The Guidelines were developed by a UNECE Task Force established by the Bureau of the Conference of European Statisticians.

II. Background and acknowledgements

5. In 2015 the Conference of European Statisticians (CES) endorsed the UNECE Guidelines on Statistical Business Registers (SBRs), which provides practical guidance and recommendations on the establishment and maintenance of SBRs.

6. The CES recognized the central role of SBRs in the production of economic statistics and the potential for using SBRs to produce business demography and entrepreneurship statistics. The CES also recognized a growing interest in entrepreneurship statistics as an area where new statistics could be developed based on the SBR and other datasets that national statistics offices have already assembled.

7. The growing importance of entrepreneurship statistics was already noted by the Bureau of the CES in 2013 when the future work on entrepreneurship statistics was discussed, based on an in-depth review prepared by OECD and Eurostat. The review highlighted areas where further work was needed to improve the production of business demography and entrepreneurship statistics, for example, by developing linkages between datasets, and longitudinal data sets.

8. With this background, in July 2016 the Bureau of the CES established a Task Force to develop guidelines for producing business demography and entrepreneurship statistics. Following on from the terms of reference for the Task Force, the guidelines describe the statistical infrastructure required to support the production of business demography and entrepreneurship statistics. More specifically, they provide:

(a) Guidance on how to develop SBRs to support the production of business demography and entrepreneurships statistics, including:

- Definitions and compilation guidelines for key demographic events.
- Suggestions of demographic statistics that could be produced directly from SBR.
- Recommendations on the requirements for the SBR to produce business demography information, longitudinally, annually and sub-annually, and by region within country.

(b) Guidance on linking the SBR with information from other data sources to produce business demography and entrepreneurship statistics.

9. The Guidelines take into account existing relevant material, in particular the *Guidelines on Statistical Business Registers* (UNECE, 2015), the *Manual on Business Demography Statistics* (Eurostat and OECD, 2007) and the *Business Registers Recommendations Manual* (Eurostat, 2010) and make use of the definitions contained in these documents.

10. The target audiences for the Guidelines are SBR statisticians in national statistics offices, experts in the production of business demography and entrepreneurship statistics as well as users of the statistics.

11. In the course of drafting the Guidelines over the period 2016 to 2018, the Task Force mainly worked through the exchange of e-mails and audio conferences. The Task Force had also two face-to-face meetings. Drafts of chapters and relevant materials were shared on a common designated UNECE wiki page.

12. During the first half-year, the Task Force agreed on the time plan of activities and worked on a detailed outline for each chapter of the Guidelines. The structure and content of the draft chapters was discussed in detail during the first face-to-face meeting that took place in Geneva over the period 13-14 June 2017. During the meeting the Task Force members noted the importance of incorporating country examples and suggested including a glossary that would be in line with the 2015 UNECE SBR Guidelines.

13. An outline of the draft Guidelines was presented to a broader audience in September 2017 during the joint UNECE, Eurostat and OECD biennial meeting of the Group of Experts on Business Registers. Meeting participants were given an opportunity to submit written comments and suggestions on the draft chapters. The comments received were taken into account in subsequent versions of the Guidelines.

14. The second face-to-face meeting of the Task Force took place on 27 September 2017 in Paris, back-to-back with the meeting of the Group of Experts on Business Registers. The draft chapters and the time plan for remaining work were discussed during this meeting.

15. From September 2017 to January 2018 the Task Force continued to work on the draft chapters and to clarify outstanding issues. With the help of an editor, a draft version of the guidelines was finalised in March 2018 and circulated to all members of the CES for written consultation.

III. Importance of measuring entrepreneurship

16. The role of entrepreneurs in stimulating economic growth has received renewed attention in the past two decades, when the extraordinary development of information and communication technologies and other advanced technologies fostered new breeds of start-ups, as well as a new attitude toward entrepreneurial activity. Here by *entrepreneurs* we mean business owners who seek to generate value through the creation of economic activity by identifying and exploiting new products, processes or markets, and by *entrepreneurship* we mean the phenomena associated with entrepreneurial activity.

17. The interest in entrepreneurship was further stimulated by the economic recession that followed the global financial crisis in 2007/2008. The notion that entrepreneurship is a crucial driver of economic growth has become part of the policy narrative on how to ensure growth and prosperity across the world, though the theoretical foundations for such an assumption are, to some extent, still under development. While the work of Schumpeter at the beginning of the 20th century had already identified entrepreneurship as a driving force for innovation and an engine for economic development, it was only in the past two decades that scholars formally proposed considering entrepreneurship as an additional factor in explaining economic growth. The assumption is that entrepreneurship is distinct from human capital and R&D, the two drivers of growth already put forward by endogenous growth theories, and that it constitutes the missing link between investment in new knowledge and economic growth. Entrepreneurs, via business creation, exploit opportunities provided by new knowledge and ideas that have not yet been discovered or commercialised by incumbent firms.

18. A second mechanism is also at work. The Schumpeterian process of "creative destruction" itself is a driver of economic growth. New firms entering the market displace obsolete firms, and the business dynamics of entry and exit contribute to productivity dynamics and eventually to economic growth.

19. Against this positive backdrop, the literature on entrepreneurship also acknowledges that entrepreneurs are a heterogeneous group, encompassing very different types. Innovative "Schumpeterian" entrepreneurs, who exploit market opportunities or innovative ideas coexist with "necessity" entrepreneurs, who start a business out of necessity as they do not have another means of generating income. While necessity entrepreneurs (also called "survival" entrepreneurs) may eventually become successful, not all entrepreneurial activities have the same, positive impact on productivity and growth. Entrepreneurship is more or less conducive to knowledge creation and diffusion, and economic growth, depending on how and where it occurs, for example in which sectors and/or locations.

20. The impact of entrepreneurship has also been associated with poverty reduction and social inclusion. Stimulating entrepreneurial activities among individuals who belong to disadvantaged or marginalised groups (because of their age, gender, ethnic characteristics, or lack of education and skills, or the geographical area where they live) can be an important vehicle for social inclusion, and it can contribute to poverty reduction by bringing into a country's active labour force segments of the population previously excluded.

21. To advance understanding of the role of entrepreneurship in economic growth and social inclusion, sound evidence on the entrepreneurial phenomenon, and its determinants and impacts, is critical. However, empirical research on entrepreneurship continues to rely to a large extent on private data sources and much less on official statistics[1].

22. Indeed, suitable and comprehensive data for the analysis of entrepreneurship are not always are found in national statistical offices (NSOs). The main challenge is that entrepreneurship data span a multitude of domains. They cross the traditional boundary between economic and social statistics, by encompassing topics such as the demography and performance of businesses, the profiles of individuals who create enterprises, the attitudes toward entrepreneurship, and the regulatory environment for setting up new businesses.

[1] At the international level, three collections of entrepreneurship data developed outside NSOs have gained popularity among analysts: the *Global Entrepreneurship Monitor (GEM)*, the *Eurobarometer on Entrepreneurship*, of the European Commission, and the World Bank's *Entrepreneurship Databases*. The first two collections rely on ad-hoc surveys of individuals, while the third mainly consists of administrative data on business registrations.

23. The Guidelines focus on economic statistics and describe concepts and methods for the production of business demography statistics to complement the wealth of business data already produced by NSOs, in particular structural business statistics. The latter have existed for many years and describe the structural characteristics of the business population, for example, business counts, employment, value added, and turnover. However, they cannot adequately account for the entrepreneurial dynamism as reflected by the creation of new businesses, their survival, growth or death. Business demography statistics, on the other hand, capture these fundamental aspects of the phenomenon of entrepreneurship.

24. The Guidelines are meant to assist NSOs and other statistical institutions in developing the production infrastructure, in particular the SBR, for compiling statistics on business demography and entrepreneurship by linking business demography data with other business and social statistics data.

25. While the primary purpose of the Guidelines is to provide advice on how to develop the SBR to facilitate the production of business demography statistics and related statistics, the Guidelines acknowledge that there are other approaches. Not having a fully developed SBR should not be considered an insuperable obstacle to production of business demography statistics.

IV. Statistics for measuring entrepreneurship

26. The topic of entrepreneurship was included as a statistical area in the Classification of Statistical Activities (CSA) only recently, in 2010.[2] Specifically, entrepreneurship is categorised under *Domain 3. Environment and multi-domain statistics*. Entrepreneurship statistics, now statistical area 3.3.7, cover "the measurement of the determinants, performance and impact of entrepreneurial activities of people and organizations". This classification acknowledges that entrepreneurship statistics are a multi-domain statistical area. For example, data on self-employment and business demography on the one hand, and data on access to finance for new and young businesses on the other, are both key elements in statistical analysis supporting the study of entrepreneurial phenomena. The classification has been inspired by the conceptual framework developed by the OECD-Eurostat Entrepreneurship Indicators Programme[3], which characterises as separate categories the *indicators of entrepreneurial determinants, performance and impact*. As illustrated in Figure 1, *entrepreneurial performance* refers to the manifestation of entrepreneurial activity, e.g., the creation of new businesses, their growth and survival, or the innovation ability of start-up and young firms. The Guidelines are concerned with, and provide guidance for, the production of indicators of entrepreneurial performance.

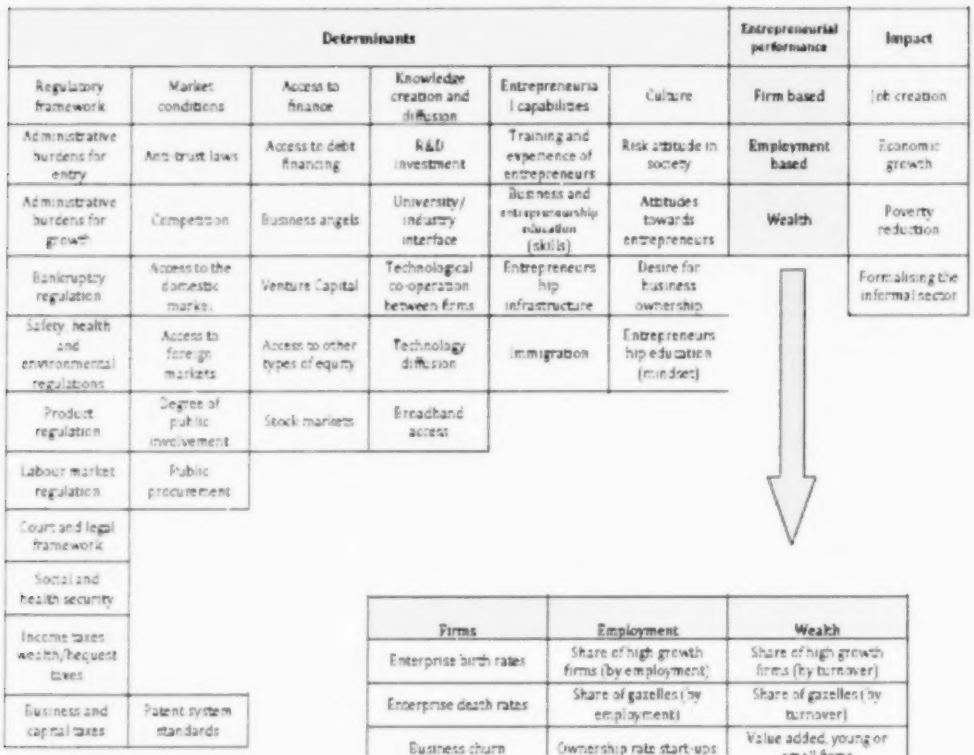
27. The Guidelines deal with measures of *entrepreneurial performance* based on business data, notably the two types described below.

[2] The Classification of Statistical Activities is an international standard for describing and categorising official statistical work by domain. See https://unstats.un.org/unsd/iaos/classification-of-International-Statistical-Activities.ashx.

[3] The OECD-Eurostat Entrepreneurship Indicators Programme (EIP), launched in 2006, develops policy-relevant and internationally-comparable indicators. The Programme produced methodological tools to structure the development and collection of indicators of entrepreneurship: "Eurostat-OECD Manual on Business Demography Statistics" (Eurostat & OECD, 2007), and the Working Papers "Defining Entrepreneurial Activity: Definitions Supporting Frameworks for Data Collection" (Ahmad & Seymour, 2008) and "A Framework for Addressing and Measuring Entrepreneurship" (Ahmad & Hoffmann, 2008).

Figure 1

Conceptual framework of the OECD-Eurostat Entrepreneurship Indicators Programme



Source: (OECD, 2017)

28. *Statistics that focus on businesses.* These primarily consist of business demography statistics measuring the births, deaths, survival and growth of businesses as well as the employment changes associated with these events. Also, by linking business demography data with other business statistics, important indicators can be generated to describe the performance of new or young enterprises versus that of mature enterprises in domains such as trade or innovation. Examples of linked business data are statistics on exports and innovation activities of young businesses.

29. *Statistics profiling businesses jointly with the business owners.* These measures of entrepreneurship rely on statistics linking data on the individuals who are business owners with data on the businesses they own. Such statistics provide the profile of an entrepreneur together with the performance of his/her business. Progress in this area has been slower reflecting the greater compilation challenges involved. Two examples are (1) statistics on

business births and deaths by gender of the business owner, and (2) innovation activity of young firms by educational attainment and employment history of the business owner.

30. The Guidelines do not cover a third category of measures of entrepreneurial performance that is characterised by having the *individual* (actual or potential entrepreneur) and not the *enterprise*, as the statistical unit. Entrepreneurship indicators on individuals include statistics on self-employment from labour force surveys and population censuses, and statistics from ad-hoc surveys of individuals on attitudes toward entrepreneurship and involvement in entrepreneurial activities. Examples are: statistics on self-employed, disaggregated by the gender and age of the self-employed; individuals' preference for self-employment over wage employment; attitudes toward the risk of business failure; and perceptions on the social and economic role of entrepreneurs. The extract in Box 1 below describes the limitations of these data as measures of entrepreneurial performance.

Box 1

Entrepreneurship and self-employment

31. The approach to measuring entrepreneurial activity has evolved significantly in the past two decades. Until the late 1990s, self-employment was the indicator most frequently used to proxy entrepreneurship. This seems a reasonable choice as self-employed persons are defined as persons who are the sole owners, or joint owners, of the unincorporated enterprises in which they work, and, therefore, they appear to represent the entrepreneur well. Also, in virtually all countries, information on self-employment is collected on a regular basis by labour force surveys and population censuses, making the use of self-employment data an easy solution for analysts in search of quantitative evidence on entrepreneurship. However, while self-employment data do contribute to the understanding of the entrepreneurial phenomenon, there are limits to their ability to represent entrepreneurship in a comprehensive and accurate way.

32. Firstly, self-employment can, at best, provide information on the population of entrepreneurs, for example, their number, their characteristics in terms of gender, age, education or work experience. It cannot offer any insight into the performance of the businesses being created, for example, their survival or growth.

33. Secondly, not all the self-employed are in fact entrepreneurs, e.g. some might have a self-employment status for reasons of convenience or – a rising phenomenon - might be "gig workers". This term describes workers engaged in flexible employment arrangements, or "gigs". While certain professions, notably in the entertainment industry, have always relied on gigs as an important source of income, the rapid popularisation of gigs has been fueled by technology and is largely associated with the rise of online platforms such as Uber and TaskRabbit that connect buyers and sellers for one-time transactions. Many gig workers use online platforms to find small jobs, sometimes completed immediately after request (essentially, on-demand), in the context of a contractual relation where any notion of entrepreneurial risk for the gig worker is absent.

34. In light of these limitations, in recent years the international statistical community has made considerable efforts to conceptualise and develop additional indicators for measuring entrepreneurial activity, in particular indicators that concentrate on businesses as the observation unit rather than individuals.

Source: (OECD, 2017)

V. Business demography and international comparability

35. Business demography statistics refer to statistics on "events like births and other creations of units, deaths and other cessations of units, and their ratios to the business population. This includes following units over time, thus gaining information on their survival or discontinuity. It also covers development over time of certain characteristics, like size, thus gaining information on the growth of individual units, or a cohort of units, by type of activity" (Eurostat & OECD, 2007). In the economic literature the term "*business dynamics*" is used to refer to the analysis of the demography of businesses and their impact on employment and productivity.[4] In these Guidelines the term business demography is used.

36. Business demography statistics can be compiled from a range of data sources, in particular SBRs, business censuses and business surveys. Research conducted in the mid-2000s on how to improve cross-country comparability of data on the creation of new businesses identified the following advantages and challenges associated with these different sources of data on business demography:

- SBRs usually provide comprehensive coverage of the population of interest and are a reliable source. In fact, data from a comprehensive, frequently updated SBR are likely to be more reliable than those from small scale surveys or studies on enterprise creation. However, the scope of the SBR, and specific threshold restrictions, can involve exclusions from the population coverage that is desired from a business demography perspective;
- Census data can be as good as SBR data, and sometimes better if they have less scope restrictions. However, the cost of running a census of businesses every year makes this approach unrealistic in any country. Data from less frequent censuses can be used, but the statistics become out of date, and their comparability of statistics across countries is difficult when their periodicity differs;
- Survey data have also been used by some countries, most notably in the Eurostat project "*Demography of Small and Medium-sized Enterprises (DOSME)*" conducted in twelve countries of Central and Eastern Europe at the time of their transition to the market economy. This approach is useful when SBRs are not sufficiently developed. It allows collection of more information on the profiles of entrepreneurs than is available from other sources. It allows collection of information on the informal economy. However, it suffers from the usual constraints set by sample size when detailed data breakdowns are required. Also, it is not possible to identify business deaths through surveys.

37. The advantages and challenges listed above also point to the use of SBRs as a most convenient source for compiling internationally comparable business demography statistics, especially if the SBRs are already subject to some form of harmonisation, i.e., if basic requirements of the registers are harmonised across countries.

38. Today, an increasing number of countries compile business demography statistics on a regular basis, although with important differences in the scope of the data collection, and mostly based on country specific approaches. The extract in Box 2 provides an overview of country practices that highlights this.

[4] Business scholars use the same term to indicate modelling methods to test the effectiveness of different policies on business outcomes and for the analysis of business strategy, for example, see (Sterman, 2000)

Box 2
Overview of country practices

39. In 2013, the United Nations Statistics Division (UNSD) conducted an "SBR Global Assessment Questionnaire Survey" aimed at collecting information on the status of the business register in all countries. The Global Survey Questionnaire was applicable not only to National Statistical Offices (NSOs) that operate a comprehensive single business register but also to those that maintain and update one or more lists of enterprises, economic census frames, or annual enterprise surveys. The list(s) of enterprises could be compiled based on multiple surveys or could be built by combining survey data with administrative data. For the purposes of the survey, all such lists were referred to as a "business register".

40. The analysis of the questionnaire findings highlighted a number of facts:

(i) In most countries, compilation of business demography statistics from SBRs started in the early 2000s or more recently. France, where the production dates to the 1970s, is an exception;

(ii) Business demography statistics produced by countries may consist simply of data on registrations and de-registrations of firms with the competent authority (as in Singapore, for example);

(iii) On the other hand, business demography statistics may be the result of implementing a methodology for identifying and handling demographic events. This is the case in OECD countries, EU member states and other developed and emerging economies where:

a *country specific* methodology is implemented, for instance, in Australia, Costa Rica, the Russian Federation and Tunisia;

an *internationally harmonized* methodology is implemented, typically based on the *Eurostat-OECD Manual on Business Demography Statistics* (Eurostat & OECD, 2007), as in EU member states, Brazil, Canada, Israel, New Zealand, Norway and Switzerland.

41. In contrast to a country specific approach is the coordinated effort conducted by the countries within the framework of the OECD-Eurostat Entrepreneurship Indicators Programme (EIP), launched in 2007. Countries that are Member States of the European Union are required to comply with mandatory regulations and to produce business demography statistics following the methodology and definitions recommended by the *Eurostat-OECD Manual on Business Demography Statistics* (Eurostat & OECD, 2007). As a result, Eurostat and the OECD have created databases of internationally comparable business demography statistics that were previously missing in the domain of official statistics.

VI. Aims of the Guidelines

42. The Guidelines aim, primarily, to support the development of the SBR as statistical production infrastructure for the compilation of business demography statistics and related indicators of entrepreneurship. Notably they address the requirements for the SBR to support the production of business demography statistics.

43. The Guidelines build on the (UNECE, 2015) Guidelines on Statistical Business Registers, and take these forward to provide guidance on developing a longitudinal SBR. They also expand the methodological explanations and provide illustrative examples of country practices for compilation of business demography statistics from SBRs.

44. The Guidelines also build on the (Eurostat & OECD, 2007) Eurostat-OECD Manual on Business Demography Statistics, which provides definitions of business demographic events (to serve as recommended international standards) and proposes a methodology to operationalise the definitions and compile business demography statistics. The Manual is supplemented but not superseded by this document, and the definitions and concepts of business demography statistics in the Guidelines are fully consistent with those in the Manual. The SBR concepts used in the Manual are based on the Eurostat Business Register Recommendations Manual and are thus consistent with the concepts and definitions in the Business Demography Manual.

45. The Guidelines complement and extend the Manual by further developing methodological aspects in relation to the compilation of statistics by linking SBR data with other data sources. The Guidelines also provide an initial discussion of issues concerning business demography at the sub-national level and concerning business demography using statistical units other than the enterprises (establishment, local units) and the need to adapt concepts accordingly. Finally, the Guidelines emphasise the effects of the compilation of business demography statistics in supporting improvements in the quality of business statistics and economic statistics.

46. In summary, the Guidelines:

(a) Deliver guidance on how to develop SBRs to support the production of business demography and entrepreneurships statistics, including:

- Definitions of key demographic events, including births and deaths;
- Suggestions for basic demographic statistics that can be produced directly from the SBR;
- Guidance regarding methods to produce business demography statistics;
- Guidance on methods for linking SBRs to other data sources (administrative/social registers and other statistical registers) to support analysis of enterprises and entrepreneurs, including gender entrepreneurship;
- Guidance on methods for linking SBRs with statistics on international activities of enterprises, such as statistics on trade, foreign-owned enterprises, and foreign affiliates.

(b) Provide examples of other approaches, not based on the SBR, to produce business demography statistics.

(c) Provide an overview of entrepreneurship indicators that can be produced and disseminated by NSOs.

VII. Overview of the Guidelines

47. The Guidelines are structured into seven chapters, three annexes and a Glossary. The following paragraphs provide short descriptions of the topics and contents of each chapter.

Chapter 2: Definitions and key concepts of business demography statistics

48. Chapter 2 presents the definitions of the key business demography concepts and related variables, based on the (Eurostat & OECD, 2007) Eurostat-OECD Manual on Business Demography Statistics and the (UNECE, 2015) Guidelines on Statistical Business Registers. The chapter provides definitions and explanations of continuity rules, and

identification of entrepreneurs, including the treatment of self-employed persons. It also discusses definitions of high-growth enterprises, including gazelles. The chapter also includes explanations on regional and sub-annual business demography, as well as the use of establishments and local units as statistical units, which is not dealt in the Eurostat-OECD Manual.

Chapter 3: SBR functionality to support production of business demography statistics

49. Chapter 3 focuses on how to develop the SBR to facilitate production of business demography statistics. The chapter aims at building the bridge between the concepts of business demography statistics and the SBR infrastructure required to support the production of business demography statistics. There are no guidelines on this topic elsewhere. Development and maintenance of longitudinal business data play a key role in the chapter, including methods to identify the demographic events and to determine whether a business has been born, survived or died. Other forms of SBR requirements are also dealt with, such as coverage of the SBR, updating frequency and the variables needed to describe demographic events.

Chapter 4: Linking SBR and business demography units with data from other sources

50. Linking business demography data with other existing data sources is an efficient way of producing statistics and increases the analytical power of the demography dataset. However, SBRs cover only a limited set of variables on the businesses and usually no information on the entrepreneur as a (natural) person. The chapter provides methods for, and guidance on, linking SBR information with other data sources. Other sources may include, for example, data on the success of enterprises (e.g., growth of employment, turnover, profit), foreign trade data, and data about the entrepreneur (e. g., sex, age, nationality, (former) employment status and education) that may be obtained from surveys, administrative registers or other statistical registers.

Chapter 5: Non-SBR based approaches to the production of business demography statistics

51. While the guidelines focus on the production of business demography statistics based on the SBR, it is recognised that, in the short run, countries might not be able to do so and production of their business demography statistics has to be based on other information. Thus, the chapter provides examples of production using data from surveys, censuses or administrative sources or based on a combination of sources.

Chapter 6: Examples of business demography and related entrepreneurship indicators

52. The chapter outlines the minimum set of business demography indicators recommended to support analysis of entrepreneurship. Examples are presented to illustrate country practices in production and dissemination of statistics on business demography.

Chapter 7: Topics for further work and research

53. The chapter includes topics for possible further work and research, based on the discussions of the Task Force. The research agenda includes: development of longitudinal databases for producing business demography statistics; linking and integrating the SBR with information from other data sources, including administrative sources; development of a statistical framework for business demography statistics; the use of different statistical units in business demography; international comparability of business demography statistics; and sub-annual (quarterly) and regional indicators.

54. The guidelines also include a Glossary with definitions and explanations of key terms and concepts; an annex presenting the development and use of Longitudinal Businesses Data in Statistics Canada; and an annex illustrating the linking of business and social statistics in Denmark to profile the entrepreneurs.